THE EFFECT OF FOUR-LANE TO THREE-LANE CONVERSION ON THE NUMBER OF CRASHES AND CRASH RATES IN IOWA ROADS

WEN LI

ALICIA CARRIQUIRY

DEPARTMENT OF STATISTICS

IOWA STATE UNIVERSITY

A research report prepared for the Iowa Department of Transportation, June, 2005

Table of contents

Acknowledgements

I would like to express my deep appreciation to my supervisor, Prof. Alicia L. Carriquiry, for her patient guidance, persistent encouragement and valuable suggestions on my work. With her kind help, I can break through the difficulties and problems that I meet all these days.

I would also like to express my deep appreciation to Dr. Michael Pawlovich who offers me a very detailed data set and materials for backgrounds of this problem, which facilitates my work. I would also like to thank all my committee members for their inspiration.

And, I wish to dedicate this thesis to my dearest parents for their ever lasting support and understanding to me in my life.

Summary

We analyze crash data collected by the Iowa Department of Transportation using Bayesian methods. The data set includes monthly crash numbers, estimated monthly traffic volumes, site length and other information collected at 30 paired sites in Iowa over more than 20 years during which an intervention experiment was set up. The intervention consisted in transforming 15 undivided road segments from four-lane to three lanes, while an additional 15 segments, thought to be comparable in terms of traffic safety-related characteristics were not converted. The main objective of this work is to find out whether the intervention reduces the number of crashes and the crash rates at the treated sites.

We fitted a hierarchical Poisson regression model with a change-point to the number of monthly crashes per mile at each of the sites. Explanatory variables in the model included estimated monthly traffic volume, time, an indicator for intervention reflecting whether the site was a "treatment" or a "control" site, and various interactions. We accounted for seasonal effects in the number of crashes at a site by including smooth trigonometric functions with three different periods to reflect the four seasons of the year. A change-point at the month and year in which the intervention was completed for treated sites was also included.

The number of crashes at a site can be thought to follow a Poisson distribution. To estimate the association between crashes and the explanatory variables, we used a log link function and added a random effect to account for overdispersion and for autocorrelation among observations obtained at the same site. We used proper but non-informative priors for all parameters in the model, and carried out all calculations using Markov chain Monte Carlo methods implemented in WinBUGS.

We evaluated the effect of the four to three-lane conversion by comparing the expected number of crashes per year per mile during the years preceding the conversion and following the conversion for treatment and control sites. We estimated this difference using the observed traffic volumes at each site and also on a per 100,000,000 vehicles. We also conducted a prospective analysis to forecast the expected number of crashes per mile at each site in the study one year, three years and five years following the four to three-lane conversion. Posterior predictive distributions of the number of crashes, the crash rate and the percent reduction in crashes per mile were obtained for each site for the months of January and June one, three and five years after completion of the intervention.

The model appears to fit the data well. We found that in most sites, the intervention was effective and reduced the number of crashes. Overall, and for the observed traffic volumes, the reduction in the expected number of crashes per year and mile at converted sites was 32.3% (31.4% to 33.5% with 95% probability) while at the control sites, the reduction was estimated to be 7.1% (5.7% to 8.2% with 95% probability). When the reduction in the expected number of crashes per year, mile and 100,000,000 AADT was computed, the estimates were 44.3% (43.9% to 44.6%) and 25.5% (24.6% to 26.0%) for converted and control sites, respectively. In both cases, the difference in the percent reduction in the expected number of crashes during the years following the conversion was significantly larger at converted sites than at

control sites, even though the number of crashes appears to decline over time at all sites.

Results indicate that the reduction in the expected number of sites per mile has a steeper negative slope at converted than at control sites. Consistent with this, the forecasted reduction in the number of crashes per year and mile during the years after completion of the conversion at converted sites is more pronounced than at control sites.

Seasonal effects on the number of crashes have been well-documented. In this dataset, we found that, as expected, the expected number of monthly crashes per mile tends to be higher during winter months than during the rest of the year. Perhaps more interestingly, we found that there is an interaction between the four to three-lane conversion and season; the reduction in the number of crashes appears to be more pronounced during months, when the weather is nice than during other times of the year, even though a reduction was estimated for the entire year. Thus, it appears that the four to three-lane conversion, while effective year-round, is particularly effective in reducing the expected number of crashes in nice weather.

Chapter 1 Introduction

1.1 Background

Conversions of four-lane undivided roads into three lanes (two through lanes and a center turn lane) are often called as "ROAD DIETS". Figure 1 (Huang et al., 2002) shows how this carried out.

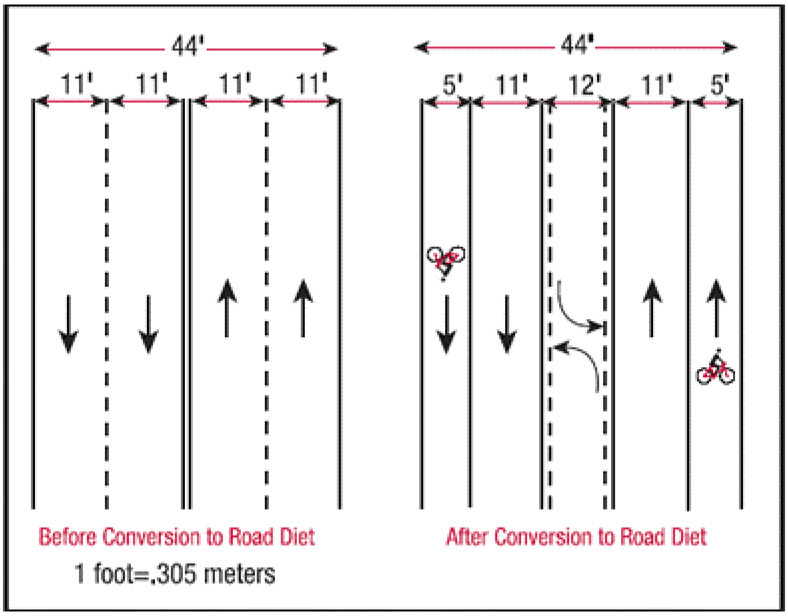


Figure 1: ROAD DIETS

In real life, the fourth lane may be used for other purpose, like being converted into bicycle lanes, sidewalks, and/or on-street parking. In other words, road diets consist in reallocating the existing space, while leaving the overall area unchanged (Huang et al., 2002)

Under most average daily traffic (ADT) conditions tested, road diets have minimal effects on vehicle capacity, because left-turning vehicles are moved into a common two-way left-turn lane. However, if road diets are used for the roads with ADTs above approximately 20,000 vehicles, there is a high probability that traffic congestion will increase to the point of diverting traffic to alternate routes (Huang et al., 2002).

As a matter of fact, road diets probably offer potential benefits to both vehicles and pedestrians. On a four-lane street, drivers change lanes to pass slower vehicles (such as vehicles stopped in the left lane waiting to make a left turn). In contrast, drivers' speeds on two-lane streets are limited by the speed of the lead vehicle. Thus, road diets may reduce vehicle speeds and vehicle interactions during lane changes, which potentially could reduce the number and severity of vehicle-to-vehicle crashes. Pedestrians may benefit because they have fewer lanes of traffic to cross, and because motor vehicles are likely to be moving more slowly (Huang et al., 2002).

Although road diet advocates enumerate these potential crash-related benefits, there has been limited research concerning such benefits. The project undertaken by the Iowa Department of Transportation is designed to help fill this gap, especially for the purpose of finding out whether four to three-lane conversions reduce the number of crashes or not. In this work, we use the term intervention to refer to road diets (the conversions of four-lane into three lanes) (Huang et al., 2002).

This report is organized as follows. In Section 1.2 we describe some of the characteristics of the study sites. Chapter 2 includes the preliminary exploratory analysis conducted on the traffic safety data collected by the Iowa DOT between 1982 and 2004 at the study sites. We use monthly crash data and monthly average daily traffic volumes (MADT) at the sites to compute the observed annual number of crashes per site and mile during the years preceding and following the intervention and also to compute the same annual totals but on a 100,000,000 AADT basis. The model that was used to estimate posterior distributions for the expected number of crashes per month and mile at each site is discussed in Section 2.2. Results are presented in Section 3. We first present and interpret the posterior distributions of model parameters. We then estimate the expected number of crashes per site and mile during the years preceding and following the intervention. The same analysis is repeated, but standardizing all sites to an equal traffic volume of 100,000,000 AADT. In Section 4, we use the model to estimate the posterior distributions of the expected number of crashes at each study site and forecast crash frequency per mile and year at each site during the months of January and June one, three and five years after completion of the four to three lane conversion. Finally, we discuss our results in Section 5. The WinBUGS program and some of the output from the WinBUGS implementation are given in the Appendix.

1.2 Data

This study evaluated road diets at several locations in Iowa. The database was constructed by the Iowa Department of Transportation. Traffic safety information on two groups of sites (a treatment and a control group) was obtained over several years before and several years after the intervention. The road diets (i.e., treatment sites) were matched with four-lane streets that were deemed to be similar to the treatment sites but did not receive the intervention (i.e., comparison/control sites). Both the fifteen pairs of treated sites and control ones are distributed over twenty-six towns and cities in Iowa: Algona, Ames, Belmond, Blue Grass, Cedar Falls, Clear Lake, Council Bluffs, Des Moines, Glenwood, Harlan, Indianola, Iowa Falls, Jefferson, Lawton, Le Mars, Manchester, Mapleton, Mason City, Merrill, Norwalk, Osceola, Oskaloosa, Rock Rapids, Sioux Center, Sioux City, and Storm Lake. The towns and cities vary greatly in population and thus in traffic volumes, but these differences were accounted for in the regression model.

Some information about the location and the population of the 30 sites is given in Table 1. Because information on site 25, the match of treated site 8 was incomplete, we used site 22 as the match for both sites 5 and site 8. Thus, site 25 in the table paired to site 8 in the table is really site 22.And, CIPOP2000 represents the census

SID	COUNTY	CITYNAME	LITERAL	CIPOP 2000	ADT 2000
1	Buena Vista	Storm Lake	Iowa 7 from Lake Ave. to Lakeshore Dr.	10076	7503
2	Cerro Gordo	Clear Lake	US 18 from N 16 st. W to N 8th St.	8161	10403
3	Cerro Gordo	Mason City	Iowa 122 from West intersection of Birch Drive to a Driveway	29172	7800
4	Clarke	Osceola	US 34 from Corporate limits on east side to where highway divides to 4 lanes on west side	4659	8172
5	Delaware	Manchester	Iowa 13 from River St. to Butler St.	5257	9400
6	Hardin	Iowa Falls	US 65 from City Limits - ? to Park Ave.	5193	10609
7	Lyon	Rock Rapids	Iowa 9 from S Greene St. to Tama St.	2573	4766
8	Mills	Glenwood	US 275 from MP 36.2 to MP 37.42	5358	6410
9	Polk	Des Moines	Beaver Ave from Urbandale Ave. to Aurora Ave.	198682	13695
10	Pottawattamie	Council Bluffs	US 6 from McKenzie Ave. west 1300 ft.	58268	11000
11	Scott	Blue Grass	Old US 61 from Oak Lane to 400' W of Terrace Drive	1169	9155
12	Sioux	Sioux Center	US 75 from 200' South of 10th St. S. to 250' North of 9th St. NW	6002	8942
13	Warren	Indianola	Iowa 92 from South R St. to Jct. of US 65/69	12998	13288
14	Woodbury	Lawton	US 20 from 100' east of Co. Rd. Eastland Ave. to 1130' West of Co. Rd. Emmet Ave.	697	9237
15	Woodbury	Sioux City	Transit Ave. from Vine Ave. to just west of Paxton St. at curve	85013	9608
18	Buena Vista	Storm Lake	Iowa 7 from Lake Ave. to Barton St	10076	8790
19	Plymouth	Le Mars	US 75 from 0.01 miles north of 3rd St NW to 0.36 miles SW of 12th St SW	9237	10880
20	Black Hawk	Cedar Falls	Green Hill Road from 0.10 miles east of IA 58 to 0.09 miles west of Cedar Heights Dr.	36145	2768
21	Greene	Jefferson	Iowa 4 from National Ave to 0.13 miles north of 250th Ave	4626	5685
22	Shelby	Harlan	Iowa 44 from US 59 to 6th St	5282	6981
23	Warren	Norwalk	Iowa 28 from 0.03 miles south of Gordon Ave to 0.04 miles south of North Ave	6884	7679
24	Wright	Belmond	US 69 from 0.38 miles north of Main St to 0.58 miles south of Main St	2560	3734
25	Shelby	Harlan	Iowa 44 from US 59 to 6th St	5282	6981
26	Polk	Des Moines	Hickman Road - 40th Place east to 0.07 miles west of W 18th St	198682	13953
27	Story	Ames	13th Street from 0.09 miles east of Stange Road to 0.07 miles west of Crescent Circle Dr.	50731	10711
28	Monona	Mapleton	Iowa 141 from 0.02 miles north of Sioux St. to 0.08 miles south of Oak St.	1322	3007
29	Kossuth	Algona	US 169 from 0.07 miles south of US 18 to 0.23 miles south of Irvington Rd.	5741	7263
30	Mahaska	Oskaloosa	Iowa 92 from 0.12 miles east of IA 432 to 0.07 miles west of Hillcrest Dr	10938	11143
31	Plymouth	Merrill	US 75 from 0.05 miles north of 2nd St to 0.18 miles north of Jackson St	754	7774
32	Woodbury	Sioux City	S. Lakeport from 4th Ave to Lincoln Way	85013	15333

Table 1: Locations of the sites and population in city as per the 2000 Census

information about the city population in the year 2000, while ADT2000 were the average daily traffic volume in the year 2000, calculated by summation of MADT in the year 2000 divided by the number of days in that year (366 days).

The Iowa Department of Transportation also provided monthly crash information at each site and estimated monthly average daily traffic (MADT) volume. Most sites were observed for 23 years, and only one site, Site 20, was observed for 21 years. Site 1- Site 15 were the sites treated, while Site 18- Site 32 were the matched controls. The details are given in Table 2 below. In the table, SID, YID, SYSTEM, DATE, YEAR denote site ID, matched pair ID, route designator, date of completion of intervention, and year of completion of intervention. One thing needed to clarify here is that LENGTH in the table below is the length of the road segment for intervention, approximately the length of the road segment for the intervention, in which road segment number of crashes was observed. And, in later parts, the length used for calculation will be the exact length of the road segment, in which number of crashes was obtained.

SID	YID	SYSTEM	ROUTE	LANES	LENGTH	COMPDATE	COMPYEAR
1	18	Iowa	7	3	1.41	1993	1993
2	19	US	18	3	1.51	May 2003	2003
3	20	Iowa	122	3	1.78	July 2001	2001
4	21	US	34	3	2.04	July 2001	2001
5	22	Iowa	13	3	0.35	July 2001	2001
6	23	US	65	3	1.23	Fall 2002	2002
7	24	Iowa	9	3	0.35	1998	1998
8	25	US	275	3	1.09	1998	1998
9	26	NA	0	3	1.19	June 1999	1999
10	27	US	6	3	0.20	April 2000	2000
11	28	NA	0	3	0.72	August 25, 1999	1999
12	29	US	75	3	1.52	1999	1999
13	30	Iowa	92	3	1.57	Summer 1999	1999
14	31	US	20	3	0.64	2000	2000
15	32	NA	0	3	0.77	2000	2000
18	1	Iowa	7	4	0.71	1993	1993
19	2	US	75	4	1.80	May 2003	2003
20	3	NA	0	4	1.80	July 2001	2001
21	4	Iowa	4	4	2.40	July 2001	2001
22	5&8	Iowa	44	4	1.20	July 2001	2001
23	6	Iowa	28	4	0.80	Fall 2002	2002
24	7	US	69	4	0.90	1998	1998
25	8	Iowa	44	4	1.20	1998	1998
26	9	NA	0	4	1.50	June 1999	1999
27	10	NA	0	4	0.33	April 2000	2000
28	11	Iowa	141	4	0.70	August 25, 1999	1999
29	12	US	169	4	2.00	1999	1999
30	13	Iowa	92	4	1.50	Summer 1999	1999
31	14	US	75	4	0.50	2000	2000
32	15	NA	0	4	1.20	2000	2000

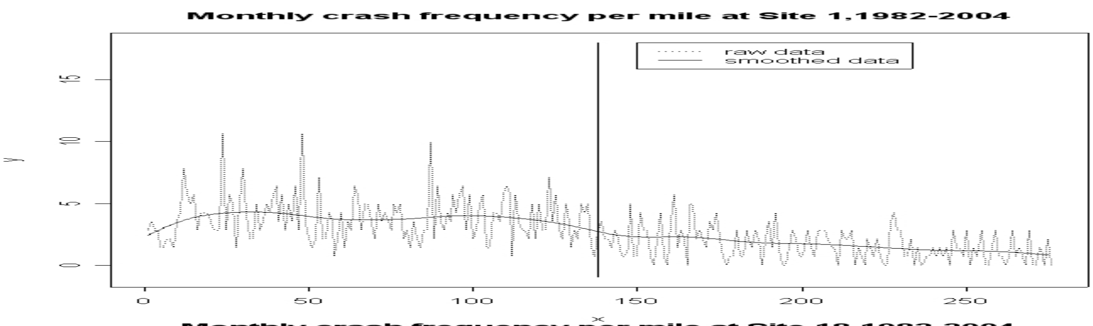
Table 2: Information recorded at the sites

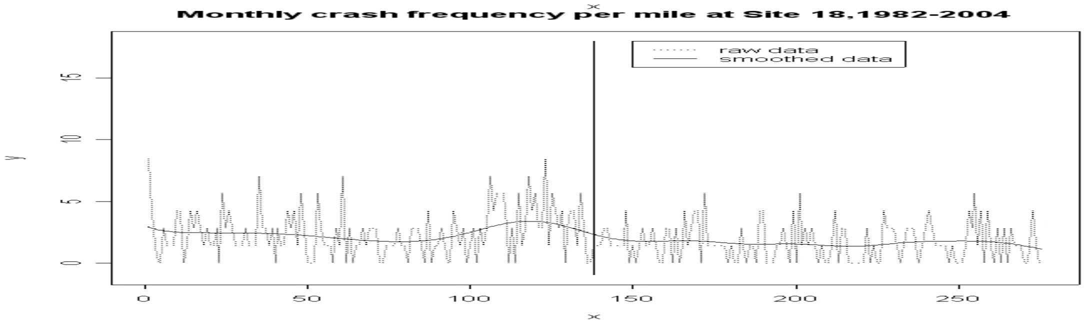
Chapter 2 Methodology

2.1 Exploratory data analysis

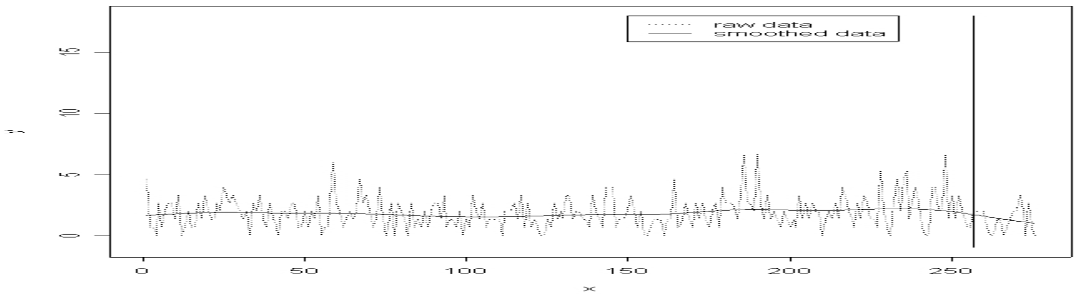
Figure 2 below shows the number of crashes per month and mile (i.e. monthly crash frequency per mile) at each of the 30 sites in the study. The y-axis is number of crashes and the x-axis is month. The vertical line in each plot marks the time at which the intervention was completed. In the plots corresponding to the matched control sites, the vertical line was placed at the month during which the intervention was completed at the corresponding treated site. The solid line in each graph is a smooth estimate of the number of crashes over time for each site. The smooth curve was obtained by fitting a non-parametric local polynomial regression with optimal bandwidth (Simonoff, J. S., 1996). We fitted the non-parametric regression model to explore the form of the Poisson regression to be fitted in later analyses. Because the length of the site varied across sites (from a low of 0.24 miles to a high of 2.53 miles) the number of monthly crashes is not strictly comparable across sites in the graphs presented in Figure 2.

Note that for most sites (treated and control) the number of crashes appears to decrease slowly over time, in an approximately linear shape, or perhaps slightly exponential fashion. Note too that seasonal effects on the number of crashes appear to be pronounced. Thus, it will be important to account for seasonality in the model for number of crashes.

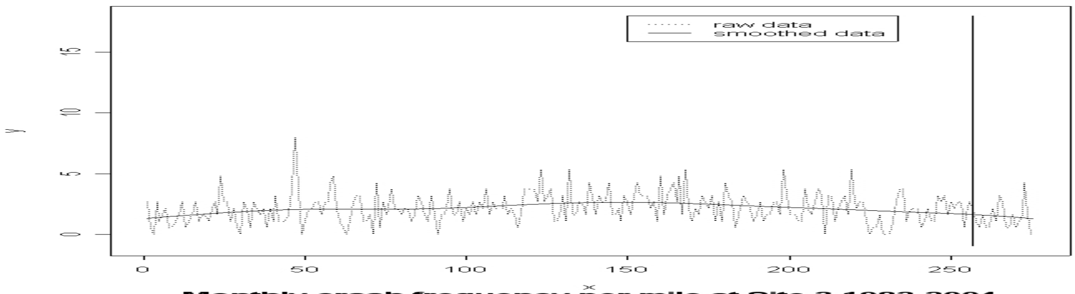




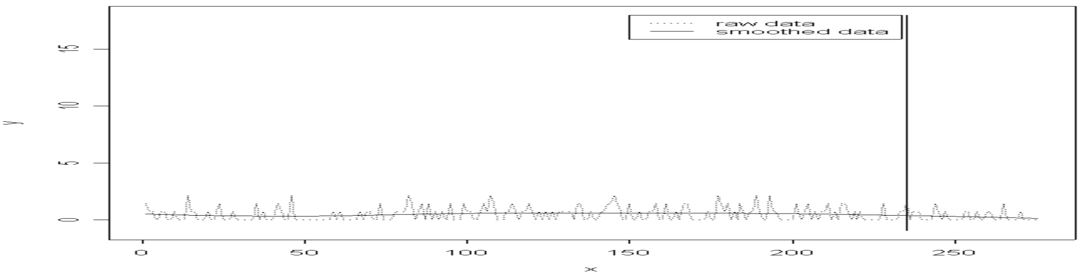
Monthly crash frequency per mile at Site 2,1982-2004



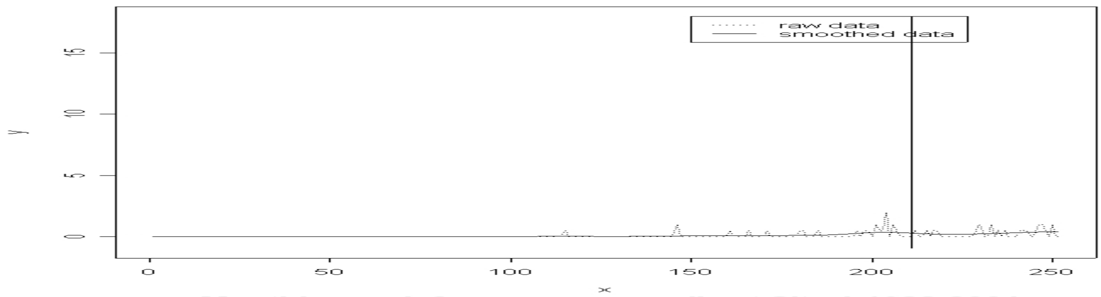
Monthly crash frequency per mile at Site 19,1982-2004



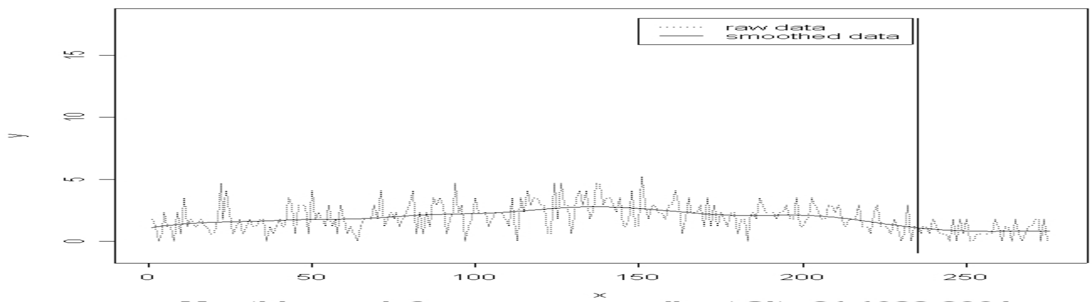
Monthly crash frequency per mile at Site 3,1982-2004



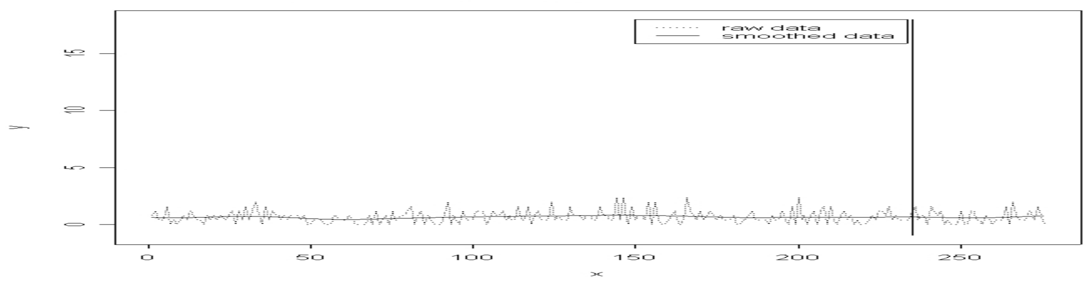
Monthly crash frequency per mile at Site 20,1984-2004

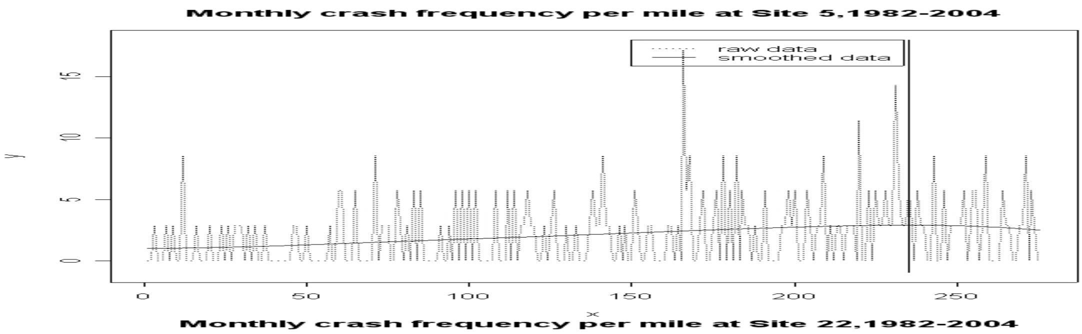


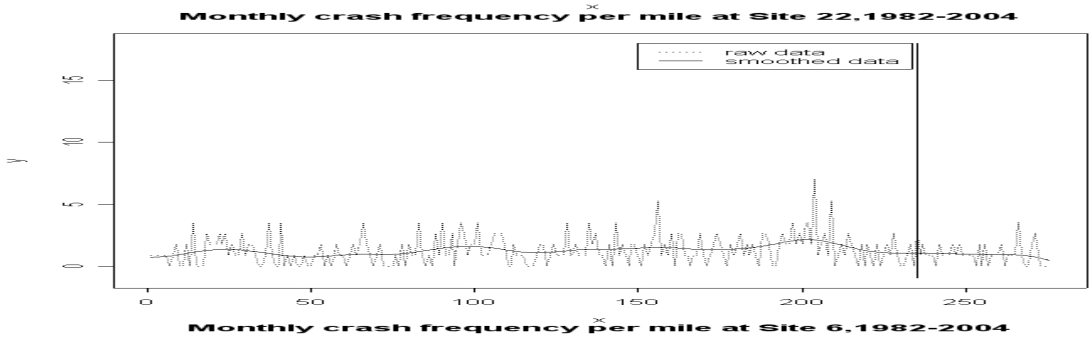
Monthly crash frequency per mile at Site 4,1982-2004

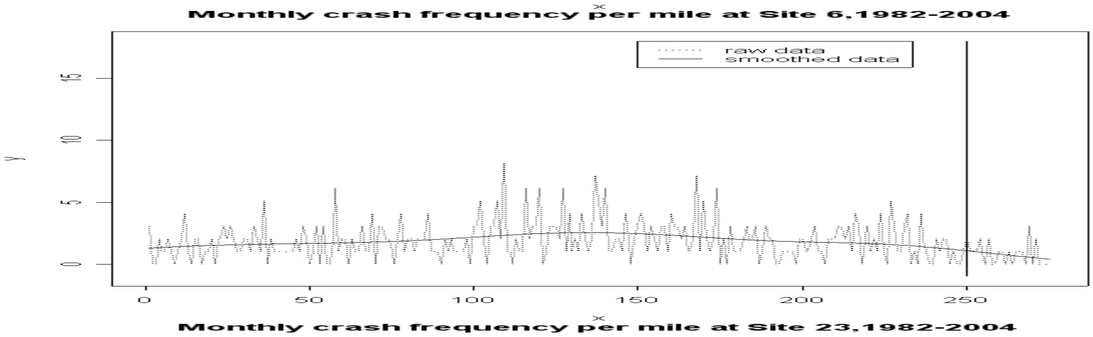


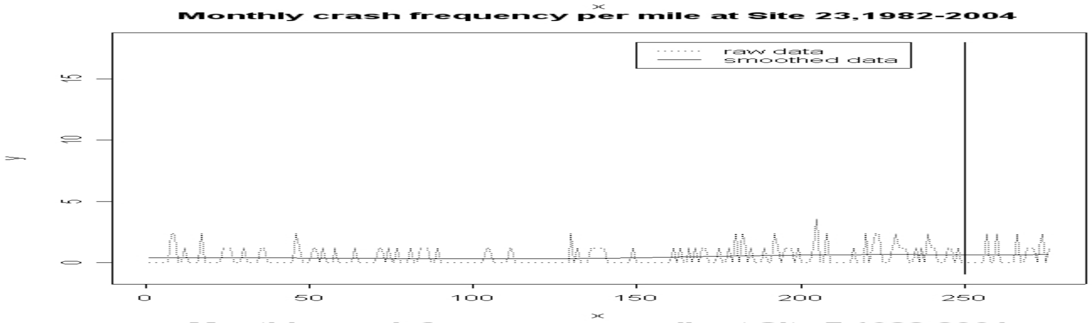
Monthly crash frequency per mile at Site 21,1982-2004

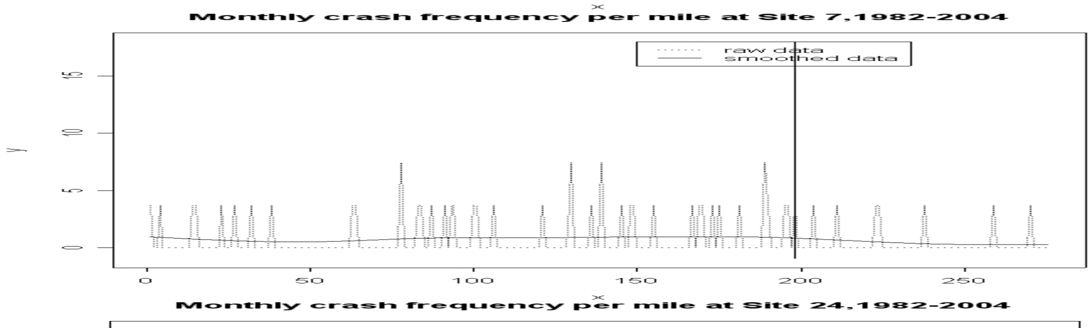


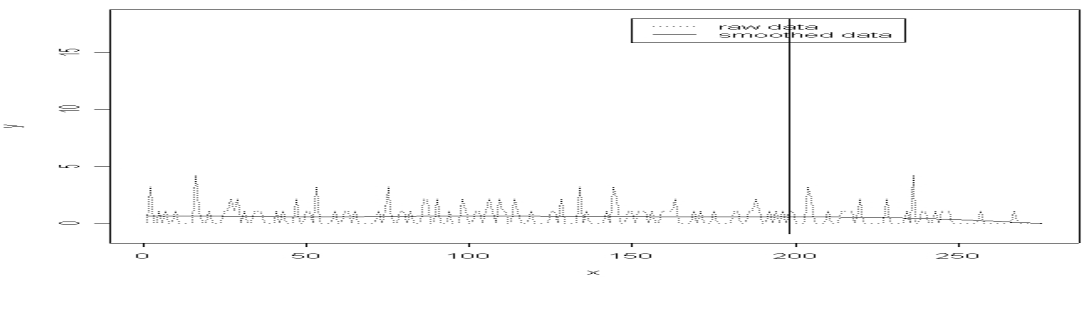

Monthly crash frequency per mile at Site 5,1982-2004
raw data
smoothed data
y
x
0
5
10
15
0
50
100
150
200
250

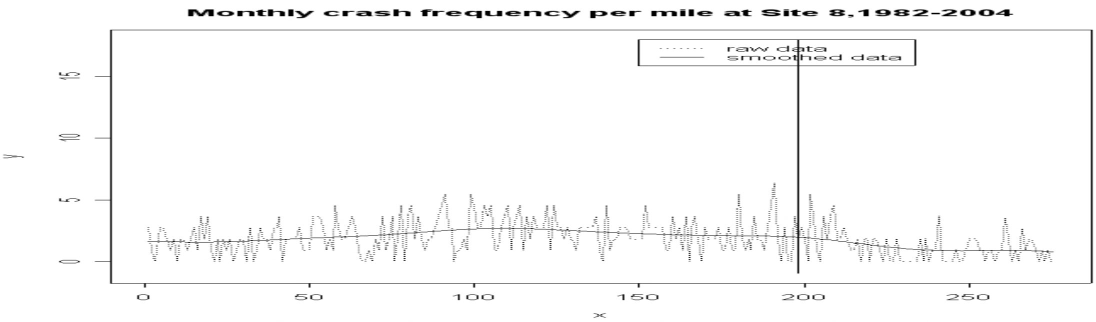

Monthly crash frequency per mile at Site 22,1982-2004
raw data
smoothed data
y
x
0
5
10
15
0
50
100
150
200
250

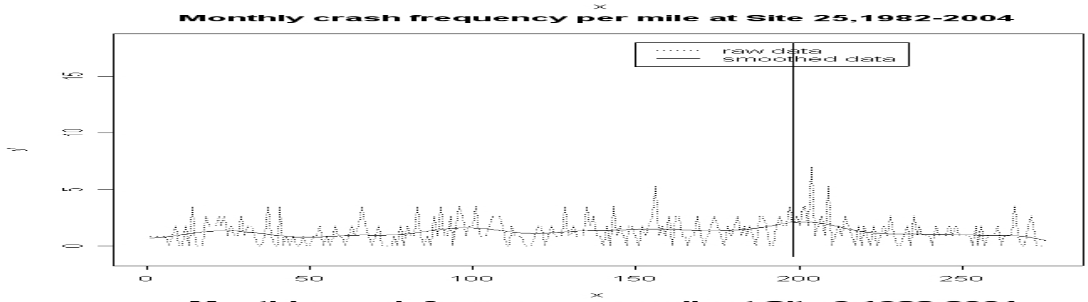

Monthly crash frequency per mile at Site 6,1982-2004
raw data
smoothed data
y
x
0
5
10
15
0
50
100
150
200
250

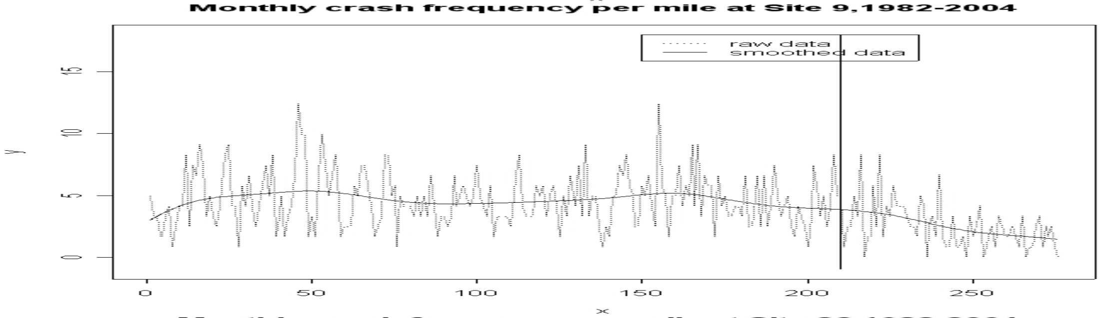

Monthly crash frequency per mile at Site 23,1982-2004
raw data
smoothed data
y
x
0
5
10
15
0
50
100
150
200
250

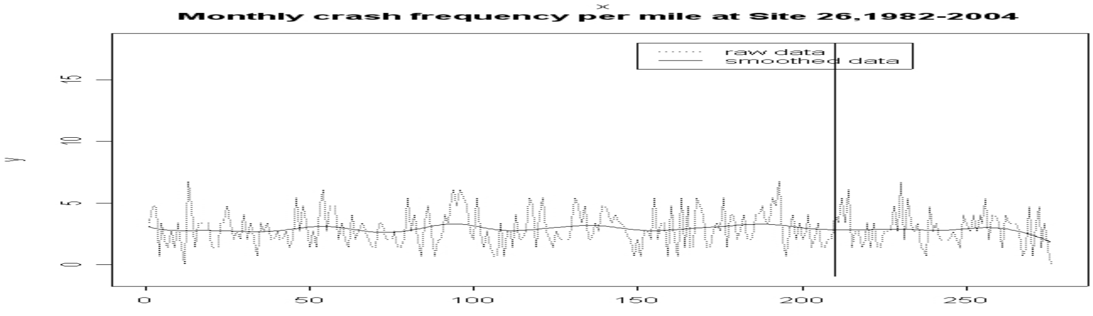

Monthly crash frequency per mile at Site 7,1982-2004
raw data
smoothed data
y
x
0
5
10
15
0
50
100
150
200
250

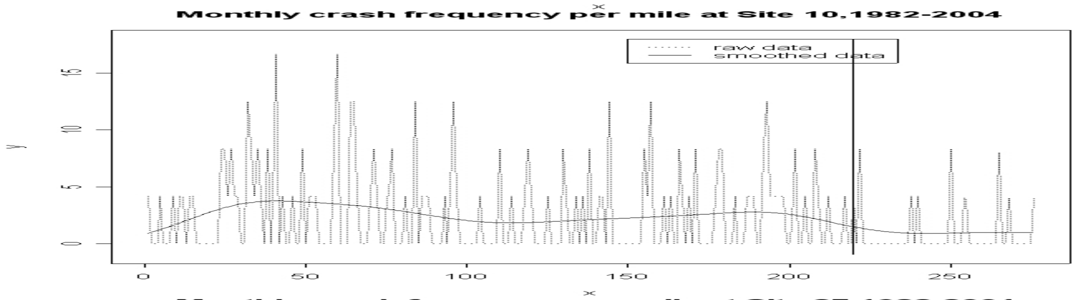

Monthly crash frequency per mile at Site 24,1982-2004
raw data
smoothed data
y
x
0
5
10
15
0
50
100
150
200
250

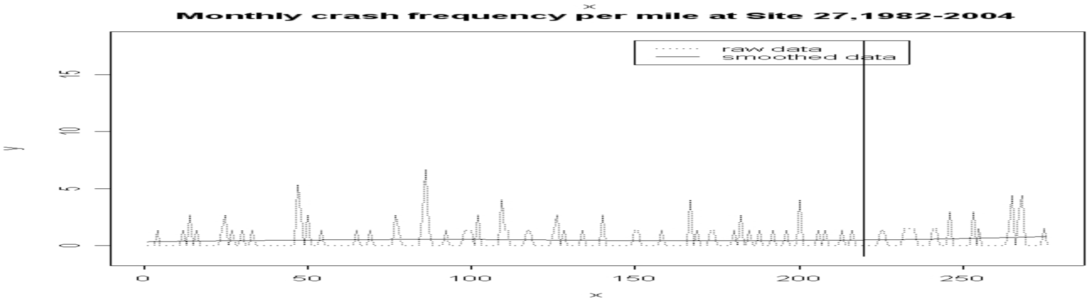

Monthly crash frequency per mile at Site 8,1982-2004
raw data
smoothed data


Monthly crash frequency per mile at Site 25,1982-2004
raw data
smoothed data

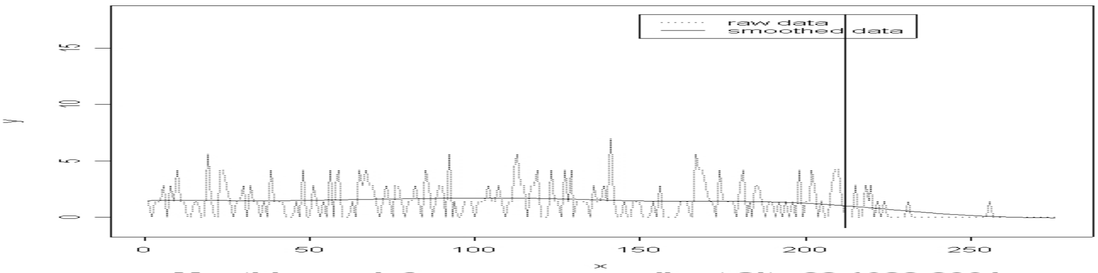

Monthly crash frequency per mile at Site 9,1982-2004
raw data
smoothed data


Monthly crash frequency per mile at Site 26,1982-2004
raw data
smoothed data

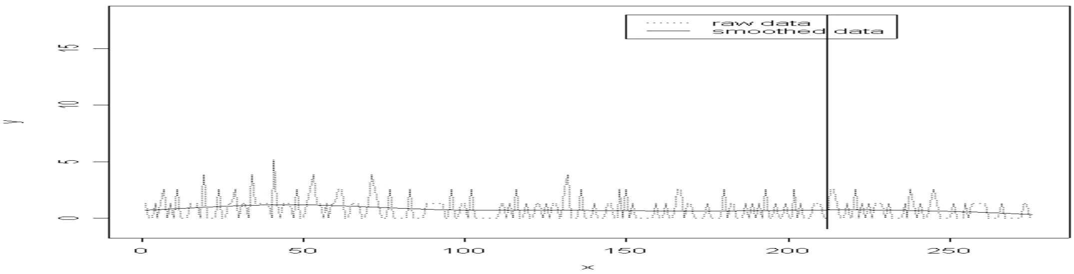

Monthly crash frequency per mile at Site 10,1982-2004
raw data
smoothed data


Monthly crash frequency per mile at Site 27,1982-2004
raw data
smoothed data

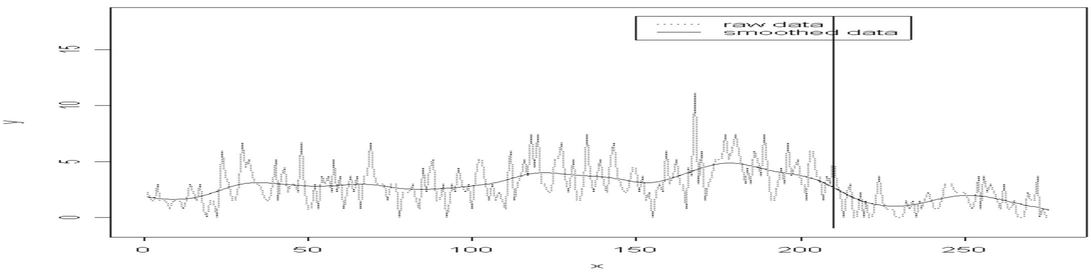
Monthly crash frequency per mile at Site 11,1982-2004



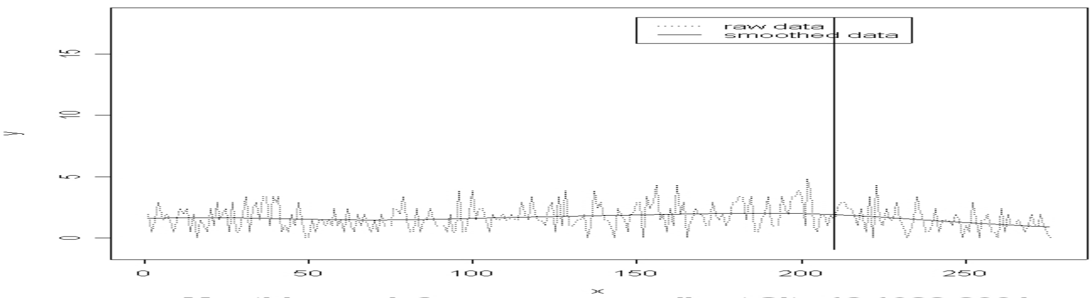
Monthly crash frequency per mile at Site 28,1982-2004



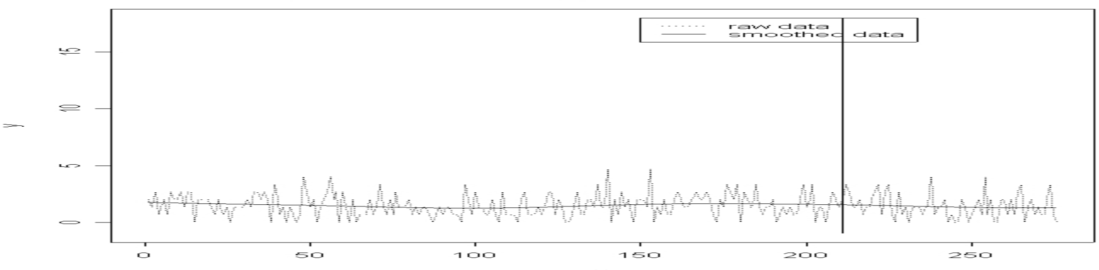
Monthly crash frequency per mile at Site 12,1982-2004



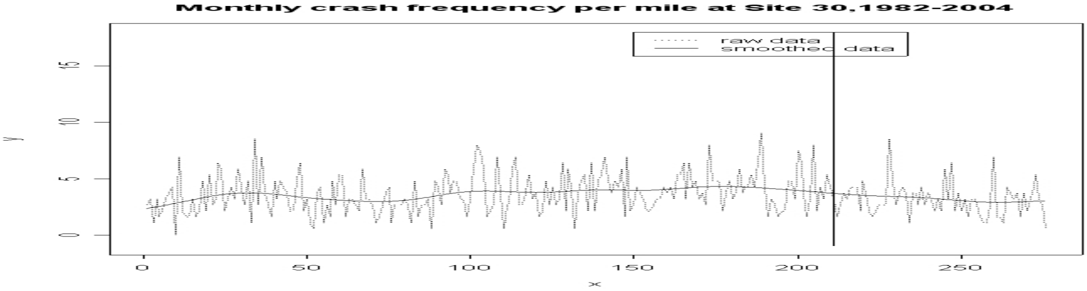
Monthly crash frequency per mile at Site 29,1982-2004



Monthly crash frequency per mile at Site 13,1982-2004



Monthly crash frequency per mile at Site 30,1982-2004



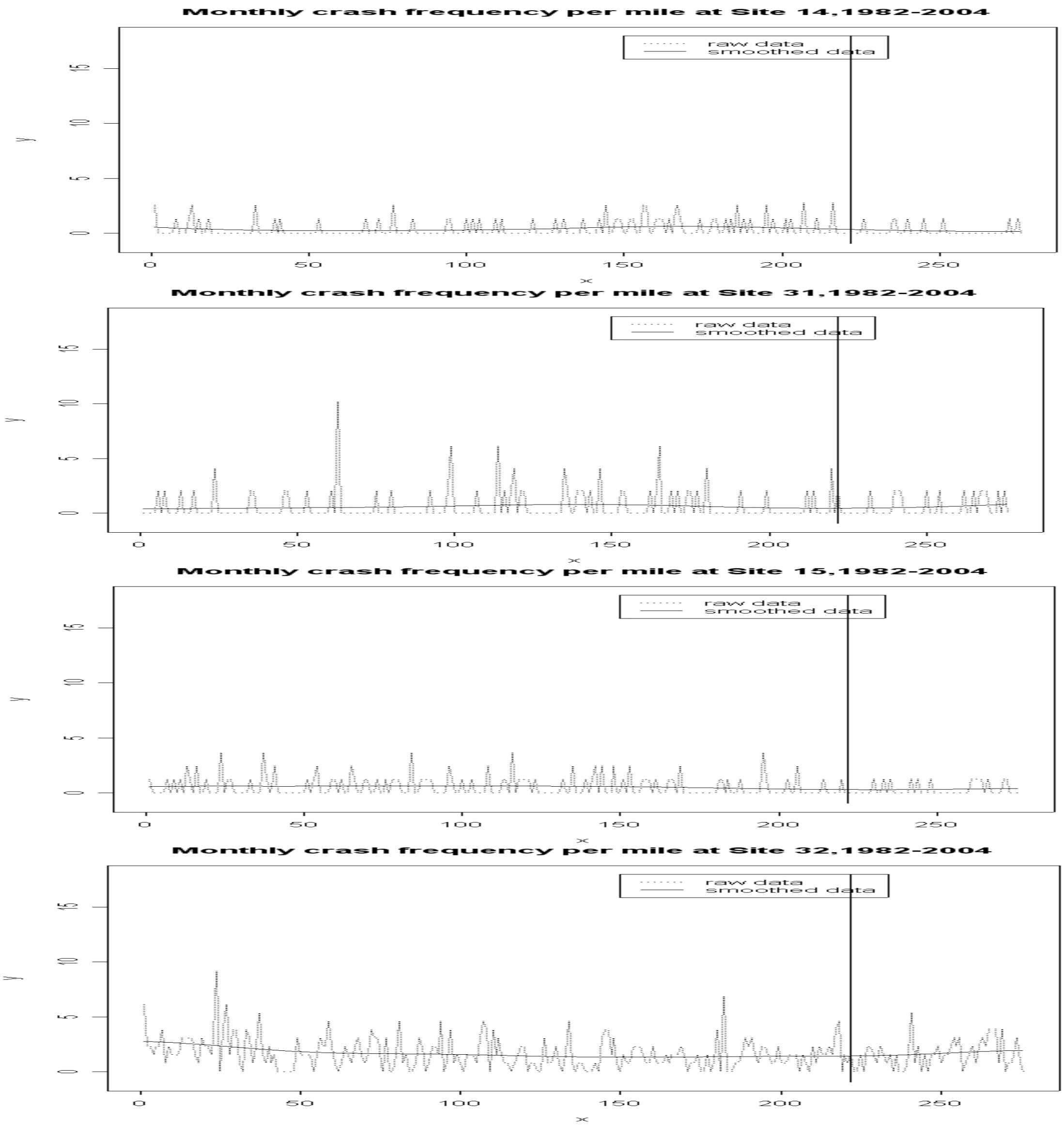


Figure 2: Observed number of monthly crash frequency per mile and smoothed estimate of number of crashes at each pair of sites.

From the graphs, we see the number of monthly crashes per mile is typically small. It appears that the number of crashes decreases after the intervention at the treated sites, as expected. However, the number of crashes per mile at the control sites also seems to decreases a bit during the same period, but not dramatically. At both treated and control sites the number of crashes per mile decreases during the entire study period even though the traffic volume at most sites increased over time. In most of the matched pairs, the evolution of crash numbers per mile before the intervention was completed is quite similar and thus we may expect to find no significant differences in number of crashes per mile in the association between the two sites and time between the two sites in a pair before the intervention is completed.

Based on the exploratory analysis above, we conclude that it is reasonable to consider a Poisson regression model with a log-link function to associate the Poisson mean to a set of covariates, since the linear model at the second (or log mean) stage corresponds to an exponential model at the level of number of crashes. To account for the clear seasonal effects we will include a set of trigonometric functions with different periods. We will also assume that the regression model on the log Poisson mean is piecewise linear, or equivalently, has a change point at the month during which the intervention was completed at the treated sites. Since it is reasonable to assume that traffic volumes have an effect on the number of crashes, we will include them as exposures in the Poisson model.

Tables 3 shows the average annual ADT for each site during the years preceding the four to three-lane conversion at treated sites and during the years following the intervention.

Site id	Average annual ADT in years before conversion	Average annual ADT in years after conversion
1	2080000	2670000
2	3440000	4390000
3	3220000	2600000
4	2810000	2960000
5	3260000	4080000
6	3130000	3850000
7	1750000	1830000
8	2250000	2300000
9	4260000	5010000
10	3280000	3970000
11	2490000	1340000
12	2990000	3670000
13	3220000	4820000
14	2330000	3230000
15	2630000	3030000
18	1840000	2810000
19	3720000	4860000
20	460000	1000000
21	1720000	2060000
22	2080000	2520000
23	2200000	2820000
24	1260000	1360000
26	4030000	5110000
27	3410000	3790000
28	930000	1090000
29	2480000	2640000
30	4240000	4220000
31	2890000	3620000
32	3080000	4480000

Table 3: Average AADT at each site during the years preceding and following the conversion from four to three lanes of sites 1 – 15.

We calculated the observed number of crashes per year and mile as well as the crash rate per year and mile (number of crashes per year and mile on a per 100,000,000 AADT basis) at each site using the monthly crash data, for the years preceding and following the four to three-lane conversion of sites 1 – 15. Those values, as well as the difference in the observed and AADT-scaled number of crashes and the percent reduction in the number of crashes during the years following the conversion are presented in Table 4 below.

SID	No. Crashes observed before conversion	No. Crashes observed after conversion	Percent reduction	Crash rate before	Crash rate after	Percent reduction in crash rate
1	46.3	19.4	58.1	2247.9	730.4	67.5
2	22.7	15.9	29.9	663.0	362.4	45.3
3	4.7	2.2	52.5	146.2	86.4	40.9
4	20.9	7.8	62.4	740.5	264.6	64.3
5	23.0	30.5	-32.0	694.5	746.9	-7.5
6	18.8	6.9	63.3	619.1	179.4	71.0
7	7.3	2.9	61.0	418.6	155.5	62.8
8	26.3	12.7	51.7	1163.5	560.5	51.8
9	57.1	28.6	50.0	1366.9	570.8	58.2
10	37.8	12.5	66.9	1176.1	314.3	73.3
11	18.3	2.5	86.3	747.1	125.2	83.2
12	34.4	14.5	57.9	1149.8	399.9	65.2
13	17.1	15.5	9.0	539.7	322.1	40.3
14	5.9	2.7	53.7	249.5	82.3	67.0
15	7.1	4.2	40.9	282.6	144.8	48.7
18	29.7	19.3	34.9	1619.8	711.9	56.0
19	27.4	18.3	33.0	738.8	377.2	48.9
20	0.6	3.1	-441.0	159.0	306.7	-92.0
21	8.2	7.8	5.4	487.3	377.0	22.6
22	14.9	10.3	31.1	716.6	407.2	43.2
23	5.6	8.1	-44.0	249.8	287.9	-15.3
24	7.4	4.6	37.1	592.8	337.3	43.1
26	34.8	33.1	4.9	877.1	647.9	26.1
27	12.8	15.9	-24.0	382.4	422.6	-10.5
28	11.0	8.0	27.3	1265.7	733.7	42.0
29	21.7	16.1	25.8	876.1	610.5	30.3
30	55.3	47.5	14.1	1300.6	1129	13.2
31	6.8	6.5	4.1	237.0	176.1	25.7
32	21.5	21.0	2.0	741.8	492.5	33.6

Table 4: Observed average number of crashes per year per mile at each site during the years preceding and following the four to three-lane conversion at sites 1 – 15, and observed percent reduction. Also, average number of crashes per year per mile per 100,000,000 vehicles (crash rate) at each site during years preceding and following conversion.

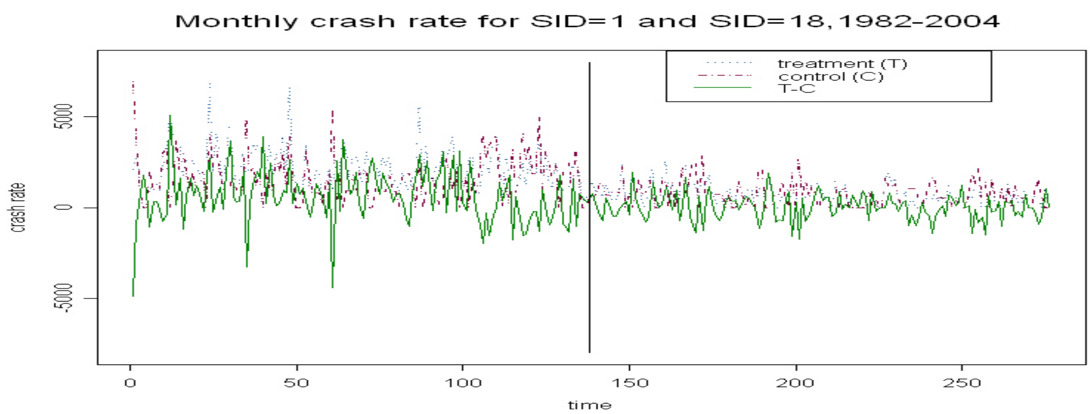
Entries in the column labeled "No. of crashes observed before intervention" is the average annual crash frequency per mile at each site, where the average was computed over all years preceding (and not including) the conversion year. Similarly, entries in the following column are the average annual frequency of crashes per mile for each site during the years following (and not including) the conversion year. The percent reduction was computed as

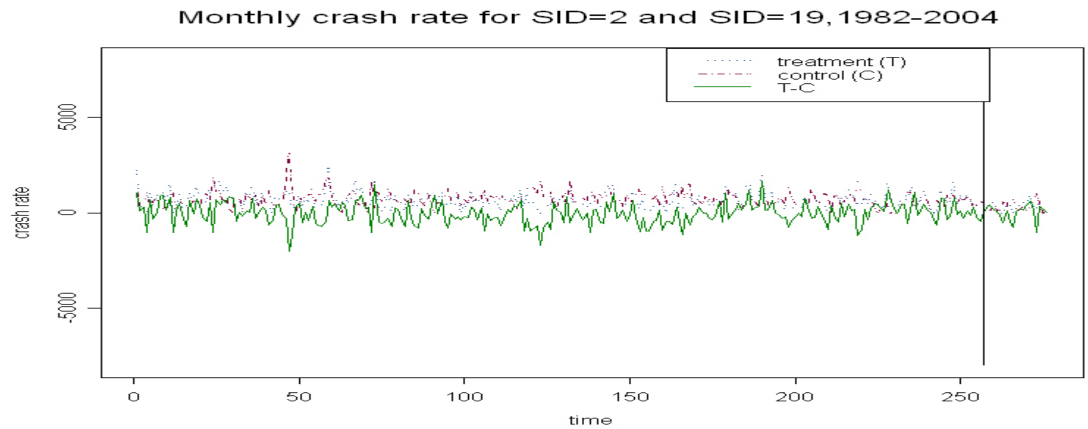
$$\textit{Percent reduction} = 100 \times (1 - \textit{Frequency after} / \textit{Frequency before})$$

Finally, entries in the columns labeled "Crash rate before" and "Crash rate after" were computed by standardizing observed crash frequencies per year and mile to a 100,000,000 AADT basis. And, we also calculate the percent reduction in crash rate

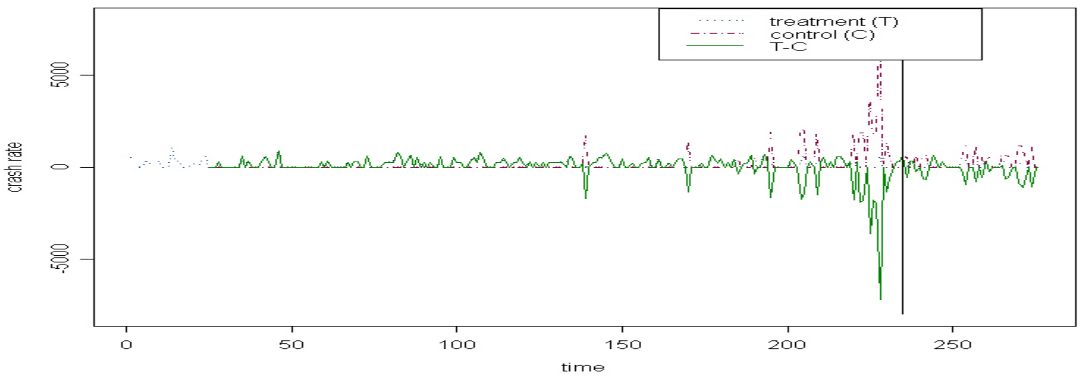
The numbers in the table suggest that the four to three-lane conversion was effective in reducing the expected number of crashes per year and mile at sites that underwent the conversion. The number of crashes appears to decline also at sites that did not get converted, but the decline seems to be less pronounced. Note that the site labeled 20 is unusual in that the number of crashes appears to increase dramatically after July of 2001, the date of completion of the conversion at Site 3, its paired site. The unexpected increase in the number of crashes at Site 20 is a result of the very small number of crashes observed at the site during the years preceding the conversion.

Figure 3 below further explores the potential differences between treated and control sites. For each pair of road segments, we drew a plot with three curves: the monthly crash rate (number of crashes per mile /traffic volume*100,000,000) for treated sites, the monthly crash rate at control sites, and the difference in monthly crash rate between the control and the treated sites (monthly crash rate of treated group – monthly crash rate of control group, solid green line). All plots are drawn at the same scale. In these graphs, the green line represents the difference in crash rate between treatment and control sites for each month. It appears that the difference is negative at most pairs of sites after the completion of the intervention than before. And, the general trend of the site-specific crash frequency, now, is clear from the plots.

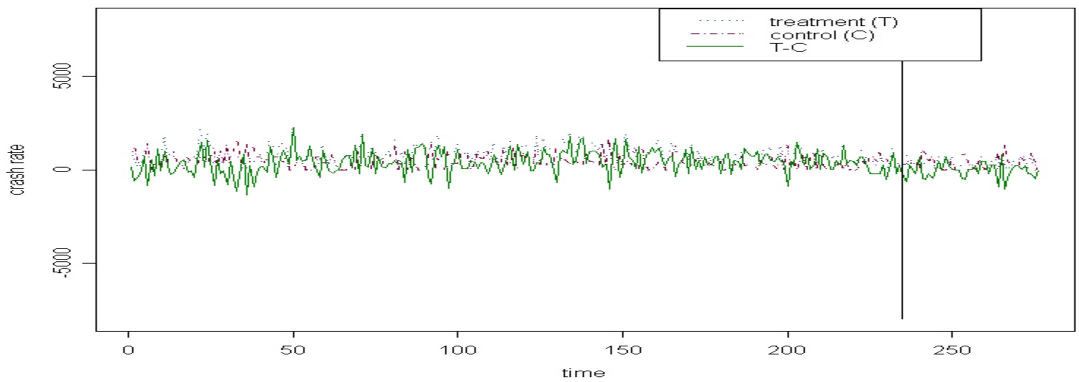




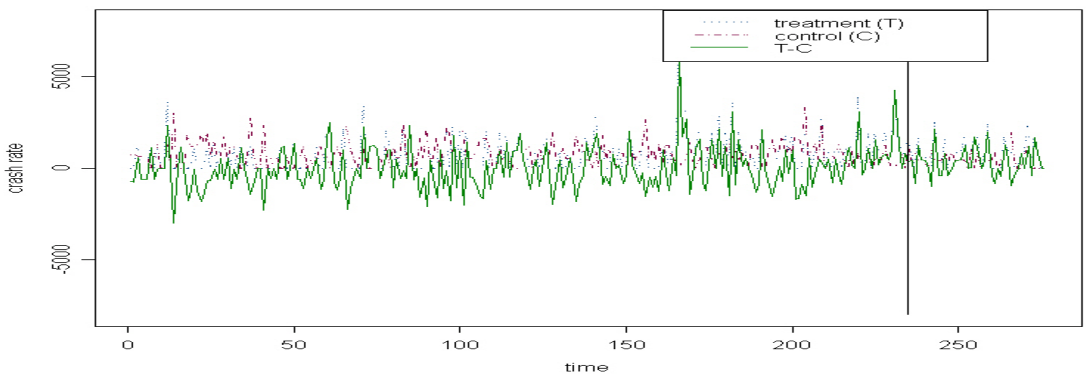
Monthly crash rate for SID=3 1982-2004, and SID=20,1984-2004



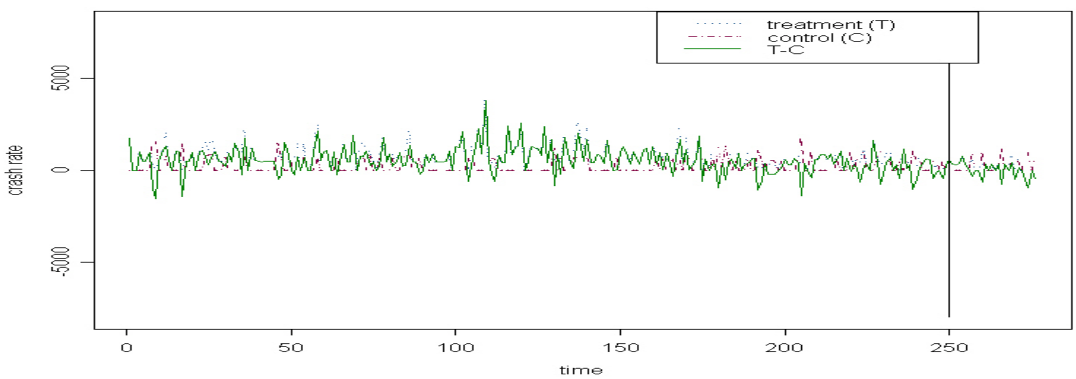
Monthly crash rate for SID=4 and SID=21,1982-2004

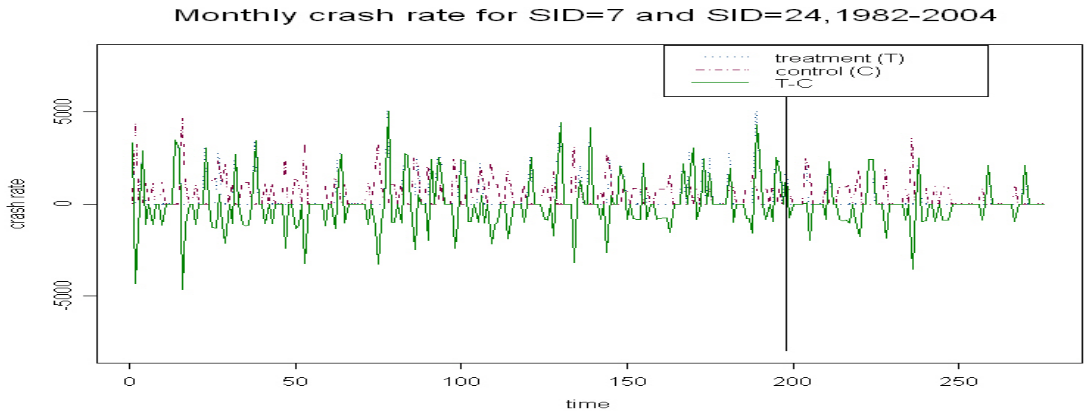


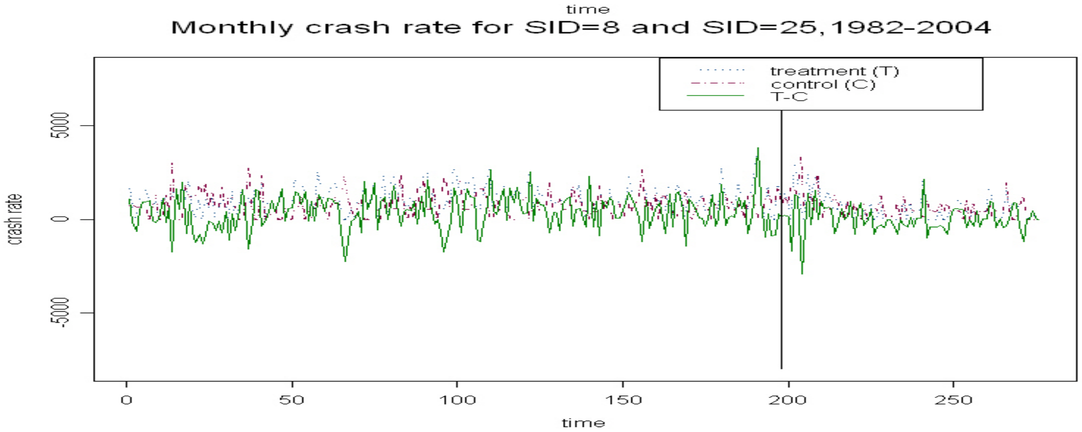
Monthly crash rate for SID=5 and SID=22,1982-2004

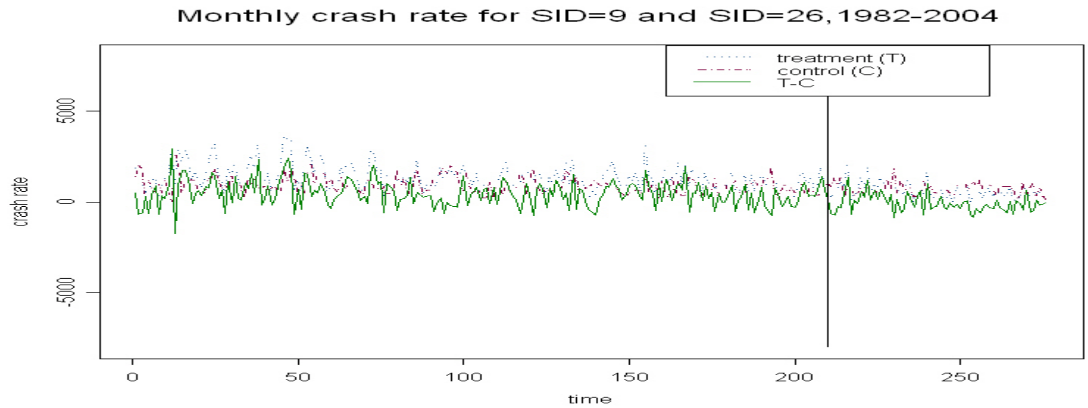


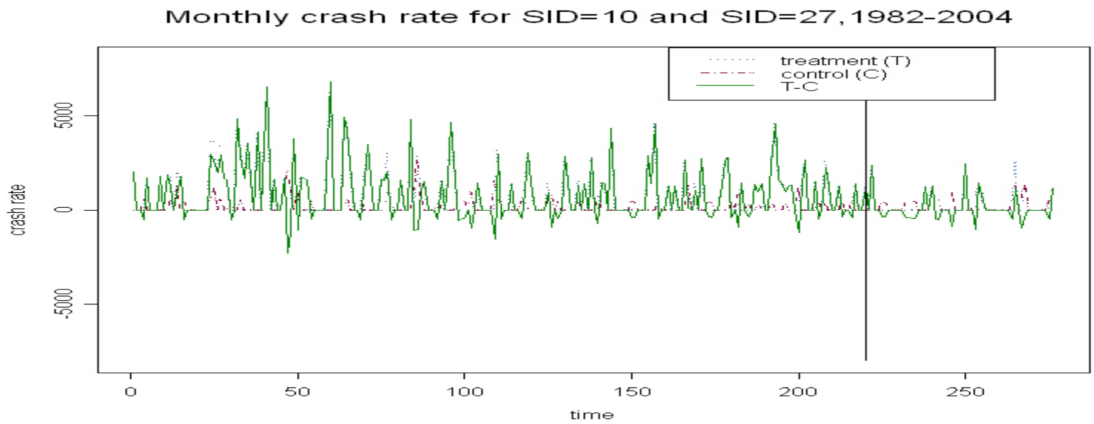
Monthly crash rate for SID=6 and SID=23,1982-2004

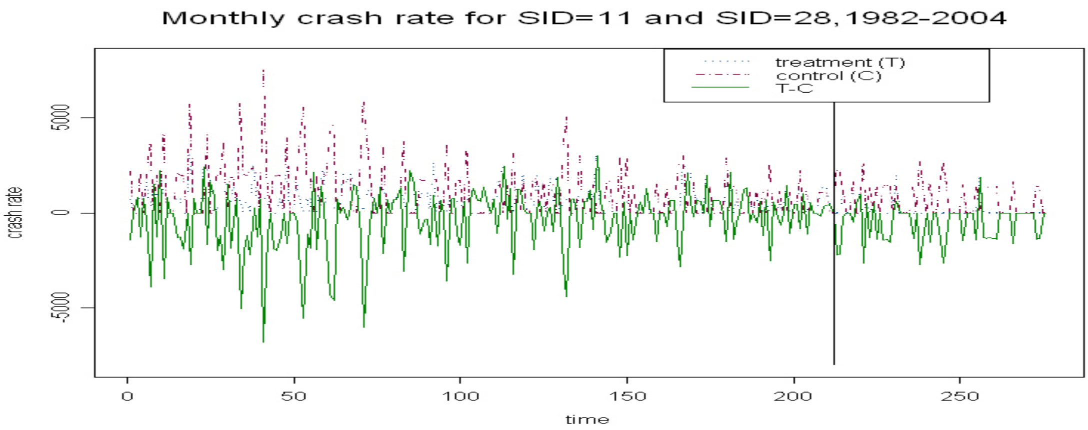


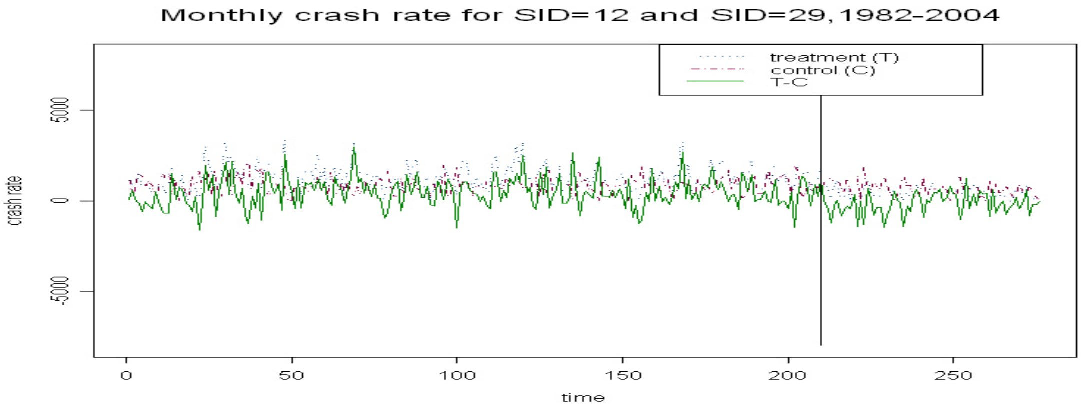

Monthly crash rate for SID=7 and SID=24, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time

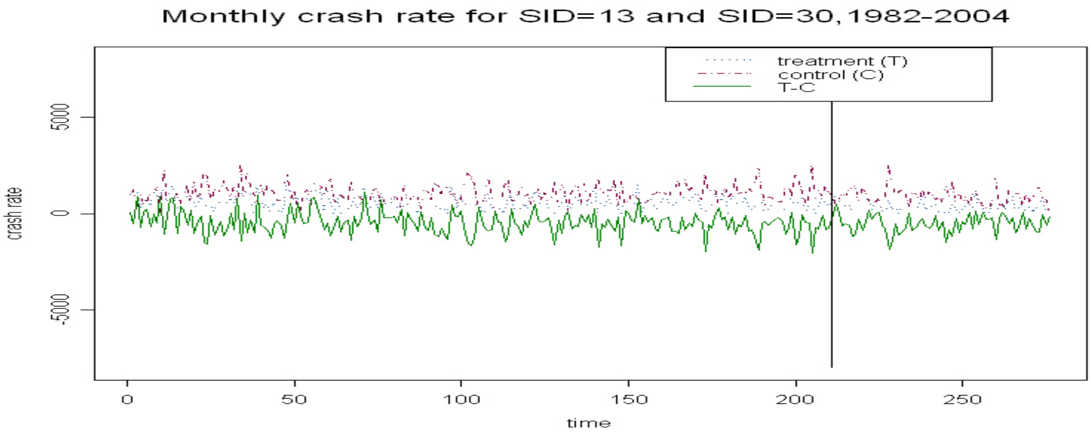

Monthly crash rate for SID=8 and SID=25, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time


Monthly crash rate for SID=9 and SID=26, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time


Monthly crash rate for SID=10 and SID=27, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time


Monthly crash rate for SID=11 and SID=28, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time


Monthly crash rate for SID=12 and SID=29, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time


Monthly crash rate for SID=13 and SID=30, 1982-2004
treatment (T)
control (C)
T-C
crash rate
5000
0
-5000
0
50
100
150
200
250
time

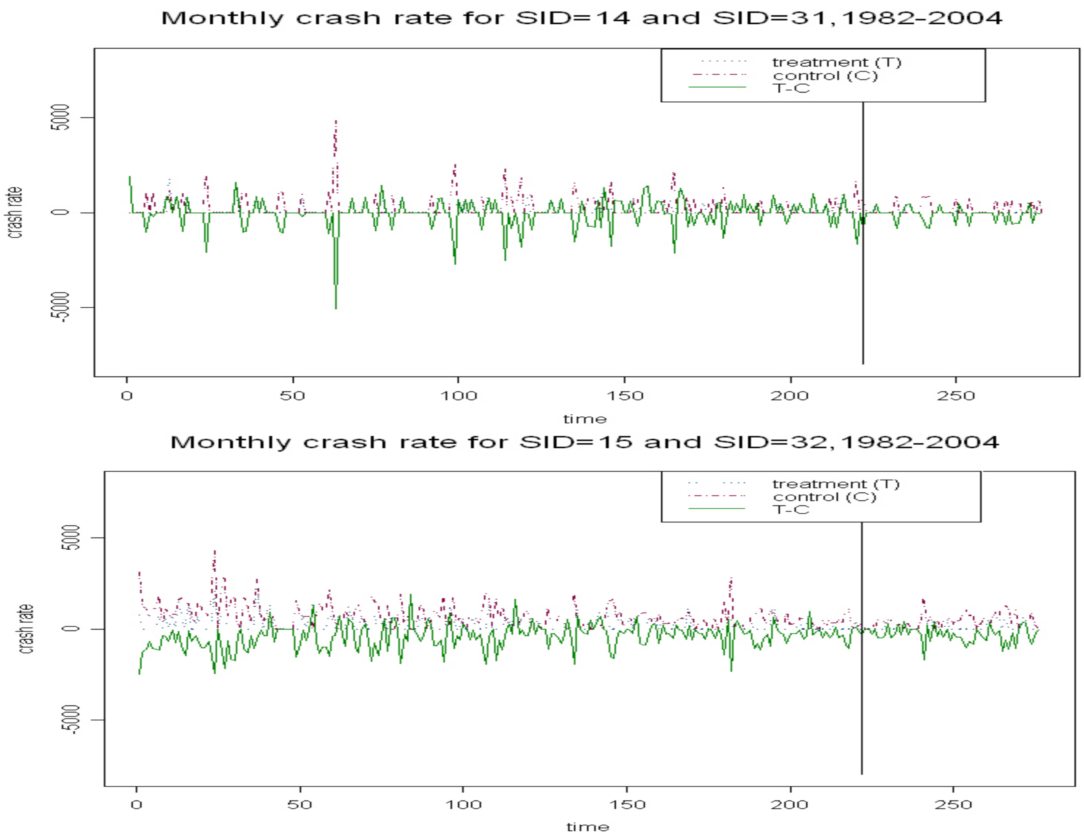


Figure 3: Monthly crash rate at treated and control sites and difference in monthly crash frequency within each pair of sites.

2.2 Formulation of the Poisson regression model

2.2.1 Model and notation

We first define notation. In the following:

i denotes site and takes on values 1 to 30,
t denotes month (or time period) and takes on values 1 to 276 (for i=17, t is from 1 to 252),
y_{it} denotes the number of monthly crashes at site *i* during time period (month) *t*,
v_{it} is the estimated monthly average daily traffic (MADT) for site *i* at time period *t*,
id_i is a random effect corresponding to site *i*,
t_{0i} denotes the time period during which the intervention is completed for treated site *i* and (fictitiously) for the corresponding matched site

$$X_{1it} := \begin{cases} 1, \text{if site } i \text{ is treated at some time} \\ 0, \text{if site } i \text{ is in control group} \end{cases},$$

$$S_{it} := \begin{cases} 1, \text{if t belongs to Winter (December, January, and February)} \\ 2, \text{if t belongs to Spring (March, April, and May)} \\ 3, \text{if t belongs to Summer (June, July, and August)} \\ 4, \text{if t belongs to Fall (September, October and November)} \end{cases},$$

and, $X_{2it} = \cos\left(\frac{2\pi \times S_{it}}{4}\right)$, $X_{3it} = \cos\left(\frac{4\pi \times S_{it}}{4}\right)$, $X_{4it} = \sin\left(\frac{2\pi \times S_{it}}{4}\right)$.

We postulate that the number of monthly crashes at a site y_{it} is a Poisson random variable with mean $\lambda_{it}v_{it}$, and divided by 1,000 just for numerical convenience. Then,

$$y_{it} \sim Poi\left(\frac{\lambda_{it}v_{it}}{1000}\right).$$

At the second level, we model the log crash rate as a piecewise linear function of the covariates defined above, such that the function is continuous at the change point. The model is

$$\log(\lambda_{it}) = \beta_1 + \beta_2 X_{1it} + \beta_3 t + \beta_4 (t - t_{0i}) I_{(t>t_{0i})} + \beta_5 X_{1t} t + \beta_6 X_{1t} (t - t_{0i}) I_{(t>t_{0i})}$$
$$+ \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i$$

where

$$id_i \sim N\left(0, \left(\tau_{bw}^2\right)^{-1}\right), \quad I_{(t>t_{0i})} = \begin{cases} 1, \text{if } t > t_{0i} \\ 0, \text{if } t \le t_{0i} \end{cases},$$

and τ_{bw}^2 is the between-site precision, defined as the inverse of the between-site variance in monthly number of crashes.

We adopt a Bayesian approach (Gelman, A., et al., 2004, and Pawlovich, 2003) for estimation and thus must choose prior distributions for all parameters at the third level of the model. We use proper, semi-conjugate but non-informative priors for two reasons. First, proper priors guarantee that the joint posterior distribution will be integrable. By letting the priors be non-informative (or almost non-informative) we let the data "speak for themselves". In this study, the number of observations available for each site, as well as the number of sites was large enough to assume that the priors will have little if any influence on the posterior distribution. The priors we chose for the regression parameters and for the precision parameter are:

$$\beta_j \sim N(0,1000), \text{for } j \text{ from 1 to 9}, \text{ and, } \tau_{bw}^2 \sim gamma(0.01,0.01).$$

A priori, none of the covariates in the model are expected to be associated to the log crash rate but prior uncertainty about this value is large since the prior variance for the regression coefficients is 1,000. The prior expected value of the precision parameter is 1. The prior variance of the precision is set at 100.

2.2.2 Interpreting the parameters of the model

Under the model, the log crash rate for a control road segment before time t_{0i} is given by:

$$\log(\lambda_{it}) = \beta_1 + \beta_3 t + \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i \ .$$

Note that at control sites during months prior to the time of completion of the intervention at the paired treated sites, the log crash rate is assumed to depend linearly on time and to be subject to seasonal variability.

During time periods after the time of intervention, when $t > t_{0i}$:

$$\log(\lambda_{it}) = \beta_1 + \beta_3 t + \beta_4 (t - t_{0i}) + \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i \ .$$

This expression can also be written as

$$\log(\lambda_{it}) = (\beta_1 - \beta_4 t_{0i}) + (\beta_3 + \beta_4)t + \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i \ ,$$

so that the control sites are allowed to have a different intercept and a different regression coefficient on time.

For a road segment receiving the intervention, the log crash rate when $t \le t_{0i}$ is given by:

$$\log(\lambda_{it}) = (\beta_1 + \beta_2) + (\beta_3 + \beta_5)t + \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i \ .$$

After the intervention, when $t > t_{0i}$:

$$\log(\lambda_{it}) = (\beta_1 + \beta_2) + (\beta_3 + \beta_5)t + (\beta_4 + \beta_6)(t - t_{0i}) + \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i \ .$$

As before, we can re-write this expression in the following form:

$$\log(\lambda_{it}) = (\beta_1 + \beta_2 - \beta_4 t_{0i} - \beta_6 t_{0i}) + (\beta_3 + \beta_4 + \beta_5 + \beta_6)t + \beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} + id_i \ .$$

If there are no differences in log crash rate between the control and the treated sites before the treatment (intervention) is completed, we would expect $\beta_2 = \beta_5 = 0$. That is, prior to the intervention we expect that both groups of sites have the same intercept and the same regression coefficient for log crash rate on time. Similarly, if there is no change in the slope of log crash rate on time before and after the treatment for the control sites, we expect that $\beta_4 = 0$.

When assessing the effect of the intervention, we compare the log crash rates between treated and control intersections after the intervention was completed. We expect that the slope of log crash rate on time is more negative after time t_0 in treated sites. That is, we expect that

$$\beta_3 + \beta_4 > \beta_3 + \beta_4 + \beta_5 + \beta_6.$$

Notice that the coefficient β_6 reflects the difference in the slope of log crash rate on time between treated and control intersections. If, as we hope, the crash rate decreases more after intervention in treated sites than in control sites, then we expect $\beta_6 < 0$.

As to X_{2it}, X_{3it}, and X_{4it}, they account for the seasonal periods which may be associated to log crash rate. Because seasonal variation is accounted for with three trigonometric functions with different periods, the regression coefficients associated to the seasonal variables are not easily interpretable. In detail, if the month *t* is in classified as a Winter month, S_{it} = 1 and then X_{2it}=0, X_{3it}=-1, and X_{4it}=1, and hence $\beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} = -\beta_8 + \beta_9$; if the month *t* is in Spring, S_{it} = 2, and then X_{2it}=-1, X_{3it}=1, and X_{4it}=0, and hence $\beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} = -\beta_7 + \beta_8$; if the month *t* is in Summer, S_{it} = 3, and then X_{2it}=0, X_{3it}=1, and X_{4it}=-1, and hence $\beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} = \beta_8 - \beta_9$; and if the month *t* is in Fall, S_{it} = 4, and then X_{2it}=1, X_{3it}=1, and X_{4it}=0, and hence $\beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} = \beta_7 + \beta_8$. We anticipate that the number of crashes will be a little higher when the weather is bad, especially in Winter, and will be lower when the weather becomes nicer, like in Summer. If so, then we expect $-\beta_8 + \beta_9 > 0$ (Winter), and $\beta_8 - \beta_9 < 0$ (Summer). We cannot easily anticipate the signs of the sums of regression coefficients associated to the Spring or Fall seasons, however.

Chapter 3 Results

3.1 Parameter estimates

We estimated the posterior distributions of the parameters in the model established above using Markov chain Monte Carlo methods and the freeware WinBUGS (Cowles, M.K., 2004). For each parameter, we ran two parallel chains over 200,000 iterations. Each chain was burned at iteration 100,001, and to avoid autocorrelation of the parameter draws, we thinned the chains, keeping every 100th draw for inference. We monitored convergence of the chains using the Gelman-Rubin statistic (Cowles, M.K., and Carlin B.P., 1996) and also checked the autocorrelation functions. After thinning, all autocorrelations were very close to zero for β_j's, for j from 3 to 9. For β_j, j =1 or 2, the autocorrelations did not decrease to almost zero even after thinning, but were not high enough to be of concern.

The posterior mean and standard deviation as well as the 2.5th, 50th and 97.5th percentiles of the posterior distributions for each of the model parameters are given in Table 5. The column entitled MC error lists the numerical (or Monte Carlo) error and we note that, as expected, the MC error is very low relative to the posterior standard deviations. Posterior distributions were approximated using 2000 draws.

Node	Mean	Std	MC error	2.5%	Median	97.5%
beta[1]	-4.755	0.2131	0.02345	-5.155	-4.762	-4.26
beta[2]	-0.2182	0.3857	0.04477	-1.05	-0.1882	0.4312
beta[3]	-0.0012	1.793E-4	4.298E-6	-0.0015	-0.0012	-8.074E-4
beta[4]	-0.00404	8.449E-4	1.909E-5	-0.0057	-0.0040	-0.002378
beta[5]	-4.51E-4	2.501E-4	7.085E-6	-9.28E-4	-4.47E-4	2.865E-5
beta[6]	-0.00915	0.001221	2.754E-5	-0.01167	-0.0091	-0.006766
beta[7]	0.03782	0.01125	2.232E-4	0.01623	0.03807	0.06018
beta[8]	-0.04558	0.007849	1.421E-4	-0.06061	-0.04574	-0.03014
beta[9]	0.08855	0.01091	2.398E-4	0.06637	0.08863	0.1093
sigma.bw	0.9553	0.1345	0.004464	0.7348	0.9383	1.231
tau.bw	1.159	0.3121	0.01025	0.6596	1.136	1.858

Table5: Summary of the posterior distributions of model parameters

From the results above, we see that β_2 and β_5 are not significantly different from zero because the 95% posterior credible sets cover zero. Thus, with high probability, we conclude that $\beta_2 = \beta_5 = 0$ as had been expected. The coefficient β_4 has a posterior distribution with mass on the negative values, which suggests that log crash rate in control intersections is lower after the month during which the intervention was completed in treated intersections. This is clearly a result unrelated to the intervention itself (as control sites never were changed from four to three lanes) and indicates that traffic safety increased at all sites in recent years.

As expected, the regression coefficient β_6 is significantly smaller than zero, since the upper bound of the 95% credible set is negative. Thus, with high probability we can conclude that the intervention was effective in that the slope of log crash rate on time is more negative in treated than in control intersections after the time of completion of the intervention. The change from four to three lanes in undivided roads in Iowa appears to increase traffic safety. Given the estimated likely set of values of β_6 we expect that a comparison of the expected crash frequency after completion of the intervention to the frequency before the intervention was implemented will show a large reduction at treated sites and a smaller reduction at control sites.

Regarding the seasonal effect on log crashes, we see that $\beta_7 > 0$, $\beta_8 < 0$ and $\beta_9 < 0$, since none of the credible sets cover zero. These results are consistent with our previous expectations. Further, we also calculated the posterior distributions of the linear combinations of the seasonal regression coefficients that were described in the previous section. We found that $\beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} = -\beta_7 + \beta_8$ which corresponds to Spring is negative, while for Fall, $\beta_7 X_{2it} + \beta_8 X_{3it} + \beta_9 X_{4it} = \beta_7 + \beta_8$ is also negative. Thus, the crash rate in Spring and Fall appears to resemble more the Summer crash rate than the Winter rate.

3.2 Interpretation of results

3.2.1 The effect of treatment (four to three lane conversion)

The main objective of this work is to determine whether road diets are effective in terms of decreasing the number of crashes at the site. If we check the posterior distribution of the coefficient of time in the control group, the treated group and the difference between them, the credible sets are all below zero. That is, the expected number of crashes at all sites decreases over time. Table 6 shows the posterior distributions of the slope of log crash rate on month after the time of completion of the intervention at treated sites, as well as the distribution of the difference in slope between treated and control sites.

Coef. for time	2.5%	50% (median)	97.5%	mean
Control (C) $(\beta_3 + \beta_4)$	-0.0067	-0.00521	-0.00371	-0.0052
Treated (Tr) $(\beta_3 + \beta_4 + \beta_5 + \beta_6)$	-0.01633	-0.01479	-0.01326	-0.01481
Diff (Tr-C) $(\beta_5 + \beta_6)$	-0.01189	-0.00958	-0.00742	-0.0096

Table 6: Posterior distributions of the slope of crash rate on month for treated and control sites, and of the difference in slope between treated and control sites.

Since the difference in the slope between treated and control sites is negative, the fitted line for log crash rate on time for the treated sites has a steeper slope than that for control sites with high probability. That is, the slope of the second segment in the piece-wise linear regression of log crash rate on time is more negative at sites which underwent the conversion than at the control sites, as is expected if the intervention is effective.

3.2.2 The effect of seasons

To more easily interpret the effect of season on crash rate, we calculated the posterior probability that the number of crashes in Spring and Fall is less than in Winter (the severest season in a year) and higher than in Summer (the mildest season in a year). Those probabilities are given in Table 7 below.

Probability	≤ Winter	≥ Summer
Spring	1.0000	0.0000
Fall	1.0000	1.0000

Table 7: Posterior probabilities that numbers of crashes in Spring and Fall are lower than in Winter and higher than in Summer.

Results suggest that the number of crashes in Spring is significantly lower than in Winter and perhaps even lower than in Summer. In Fall, the number of crashes is lower than in Winter, but higher than in Summer. To better understand the effect of season, we calculated the 95% credible sets for the appropriate linear combinations of regression coefficients. Results are presented in Table 8 below.

	2.5%	50% (median)	97.5%	mean
Winter $(-\beta_8+\beta_9)$	0.10831	0.13405	0.15923	0.134128
Spring $(\beta_8-\beta_9)$	-0.15923	-0.13408	-0.10831	-0.13413
Summer $(-\beta_7+\beta_8)$	-0.10963	-0.08383	-0.05606	-0.08341
Fall $(\beta_7+\beta_8)$	-0.03491	-0.00745	0.01852	-0.00776

Table 8: 95% posterior credible sets for linear combinations of regression coefficients representing the effect of seasons.

The credible set for the linear combination of parameters representing Winter is above zero, while the one for Summer is below zero, which is consistent with intuition. That is, the log crash rate tends to be higher in Winter and lower in Summer.

However, the season associated to the lowest crash rates is Spring. Notice that the credible set of the linear combination of regression coefficients representing Spring is below zero, and that all quantiles and posterior mean are lower than those corresponding to Summer. One possible explanation for this is that drivers, cautious after the difficult driving during winter months experience fewer crashes as weather (and thus traveling conditions) improves. Similarly, as the weather gets worse in Fall, the number of crashes due to drivers used to Summer conditions tends to increase.

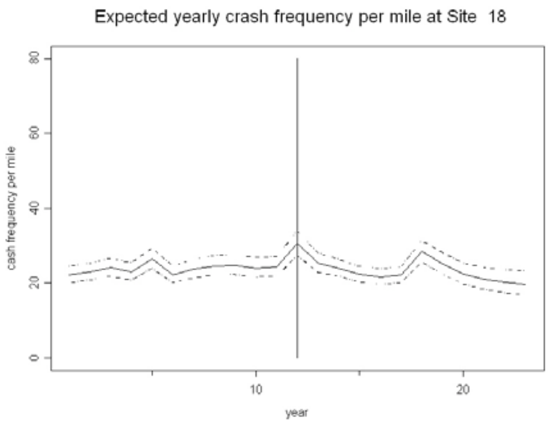
Chapter 4 Expected crash frequencies at study sites

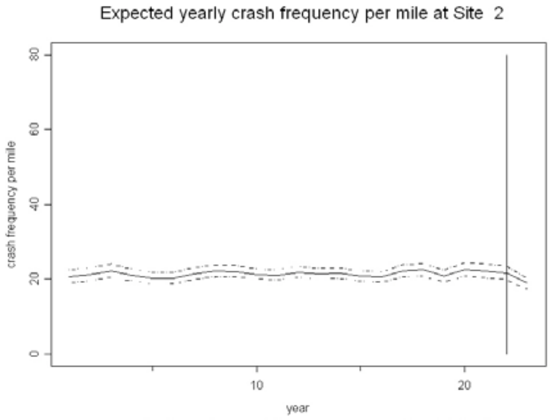
4.1 A comparison of expected crash frequencies after and before conversion

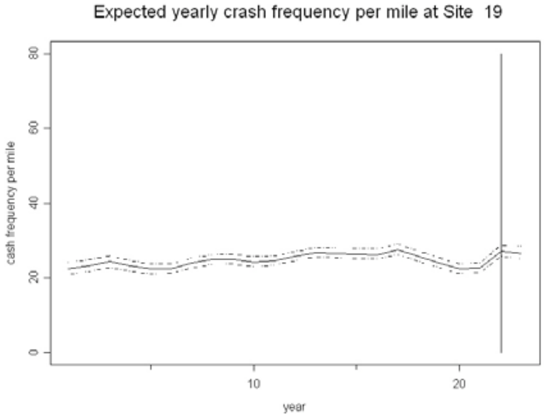
The main objective of this work was to assess whether the four to three lane conversion at the 15 sites was effective in increasing safety. The observed crash frequencies presented in Table 4 suggest that in fact there was a more pronounced reduction in crash frequency at converted sites than at sites that were not modified, even though traffic volumes appear to have increased at most sites.

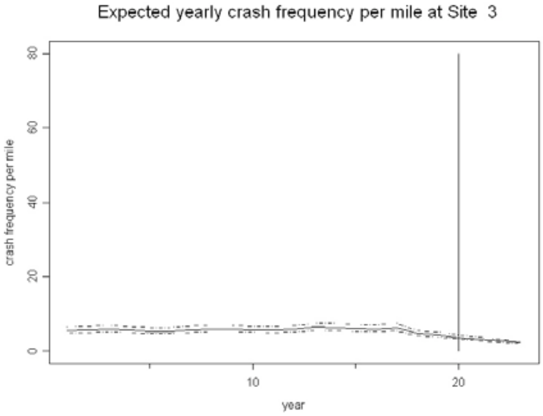
In this section, we estimate the posterior distributions of the average expected number of crashes per year and mile for each of the study sites during the years preceding and also following the conversion at sites 1-15. To quantify the difference in crash frequency during the "After" and "Before" periods, we computed the posterior distribution of annual frequency per site and mile and also the posterior distribution of the percent reduction in crash frequency for each site. The overall reduction was computed from the average, over the treated and over the control sites, of the site-specific "Before" and "After" crash frequencies per mile.

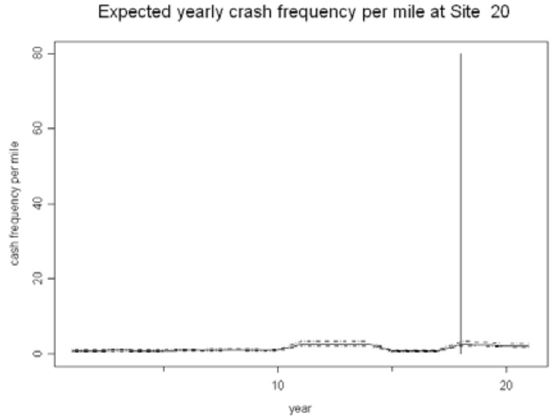
Figure 4 below shows, for each site in the study, the posterior mean of the expected yearly number of crashes per mile (i.e. expected yearly crash frequency per mile) and the 2.5th and 97.5th percentiles of the posterior distribution of crash frequency. The solid vertical lines on each plot mark the year of completion of the intervention at the treated sites.

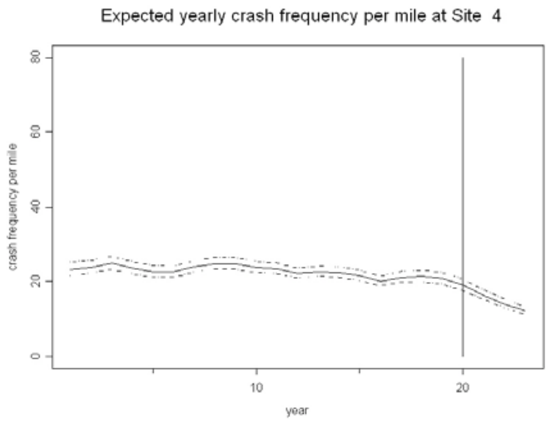


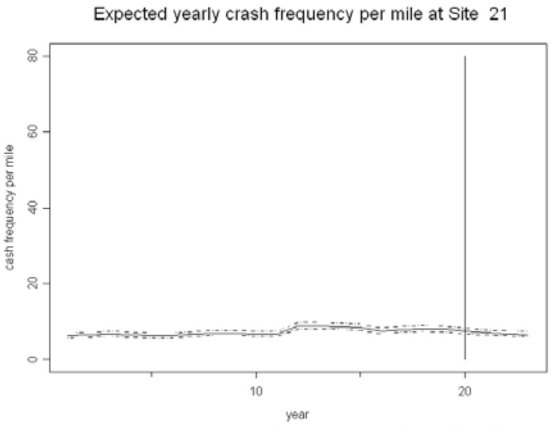

Expected yearly crash frequency per mile at Site 2
crash frequency per mile
year

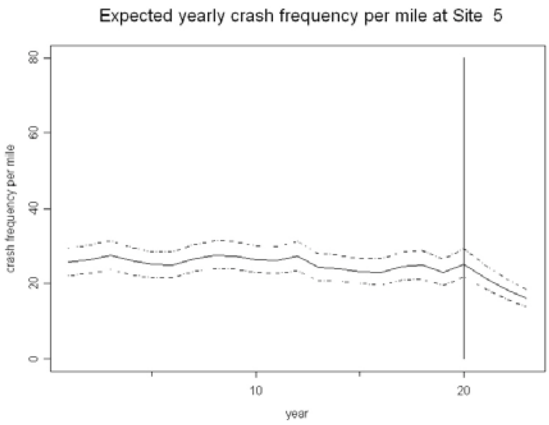

Expected yearly crash frequency per mile at Site 19
cash frequency per mile
year

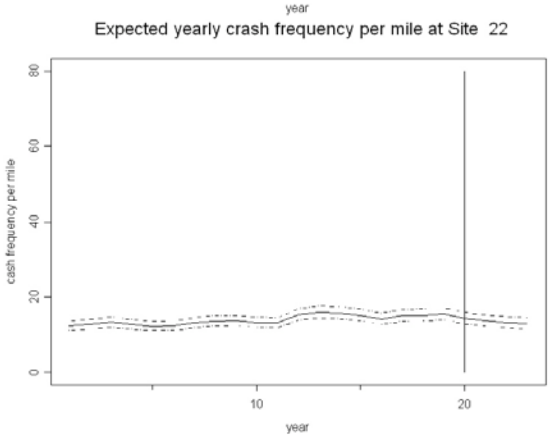

Expected yearly crash frequency per mile at Site 3
crash frequency per mile
year


Expected yearly crash frequency per mile at Site 20
cash frequency per mile
year

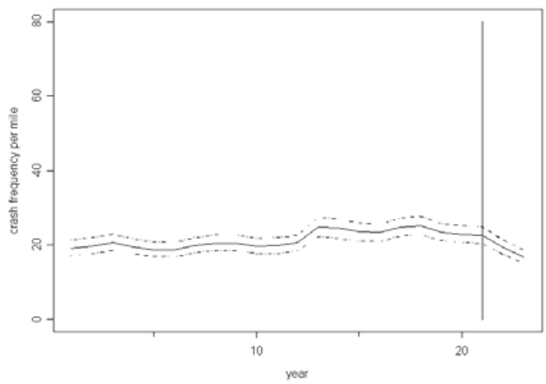

Expected yearly crash frequency per mile at Site 4
crash frequency per mile
year


Expected yearly crash frequency per mile at Site 21
crash frequency per mile
year

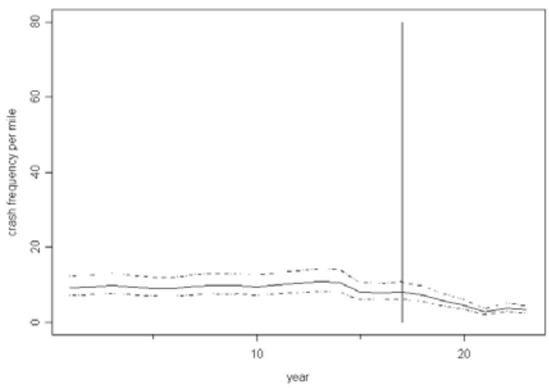

Expected yearly crash frequency per mile at Site 5
crash frequency per mile
year


Expected yearly crash frequency per mile at Site 22
cash frequency per mile
year

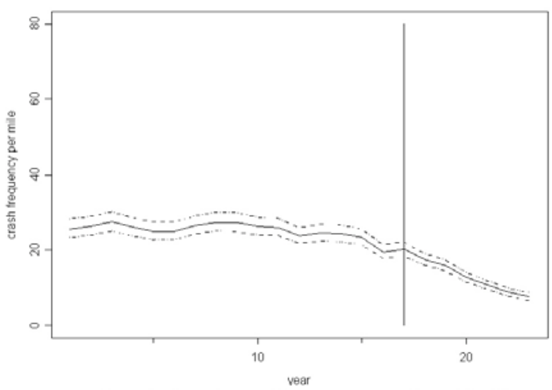
Expected yearly crash frequency per mile at Site 6



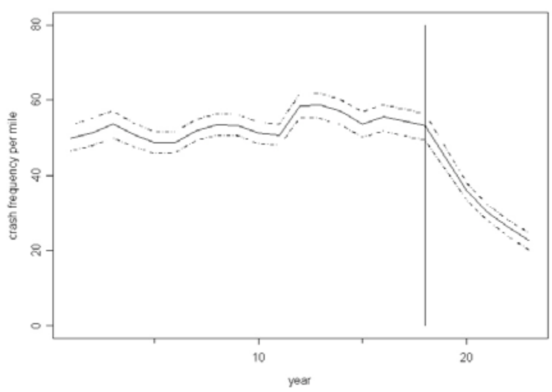
Expected yearly crash frequency per mile at Site 7



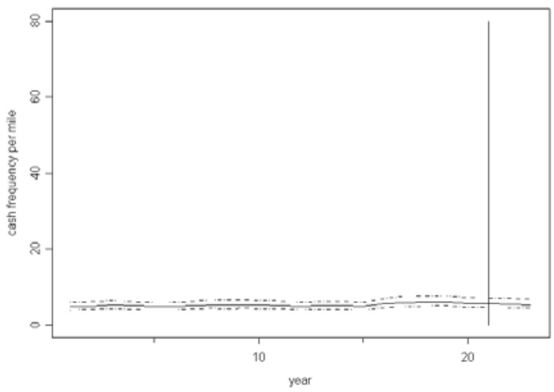
Expected yearly crash frequency per mile at Site 8



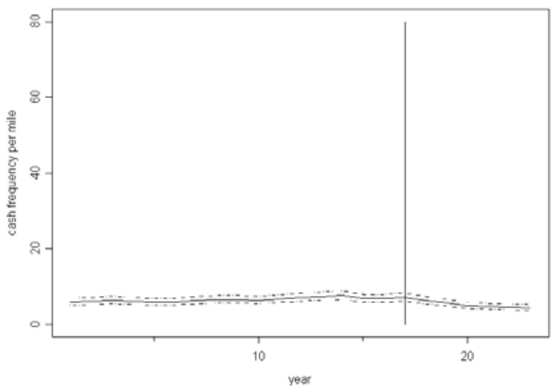
Expected yearly crash frequency per mile at Site 9



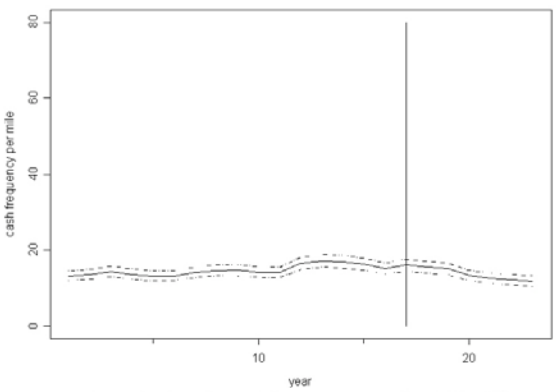
Expected yearly crash frequency per mile at Site 23



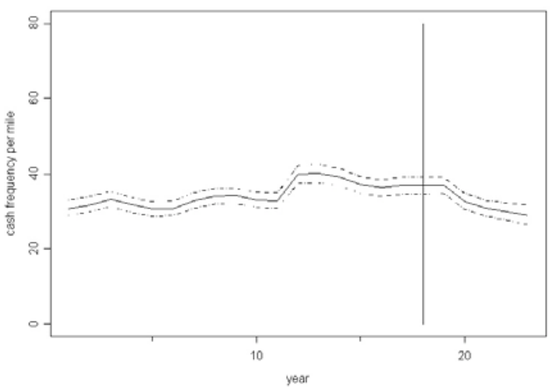
Expected yearly crash frequency per mile at Site 24



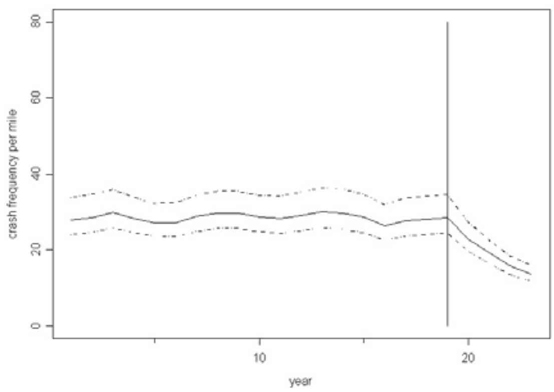
Expected yearly crash frequency per mile at Site 25



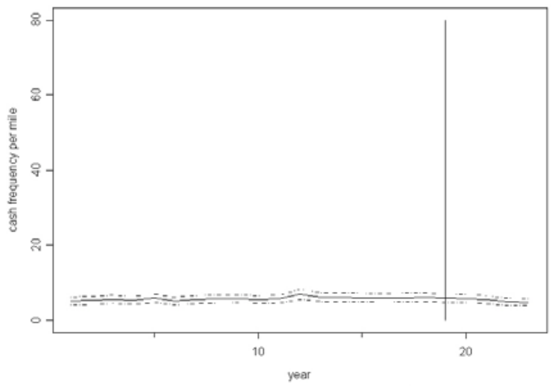
Expected yearly crash frequency per mile at Site 26



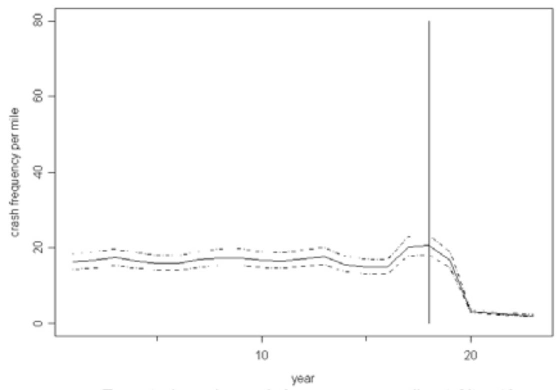
Expected yearly crash frequency per mile at Site 10



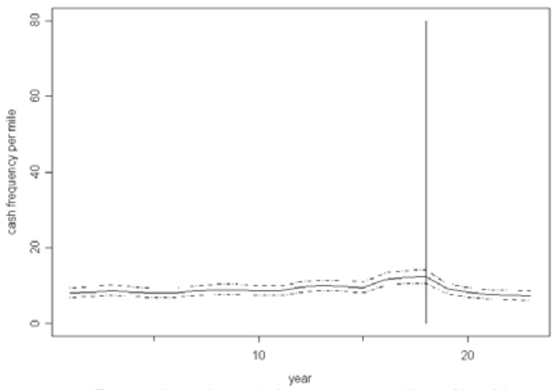
Expected yearly crash frequency per mile at Site 27



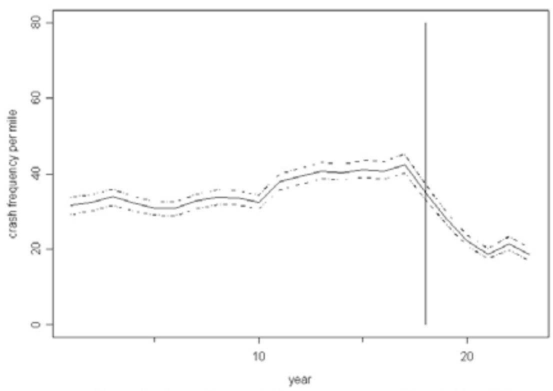
Expected yearly crash frequency per mile at Site 11



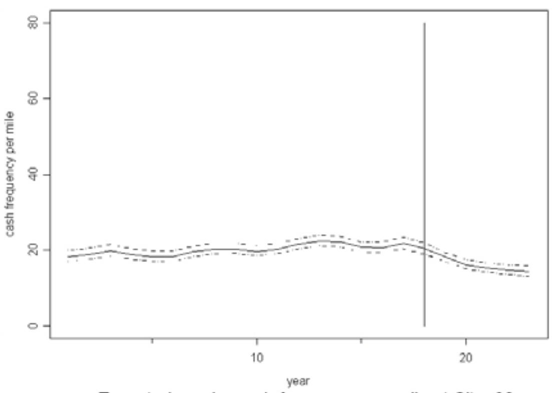
Expected yearly crash frequency per mile at Site 28



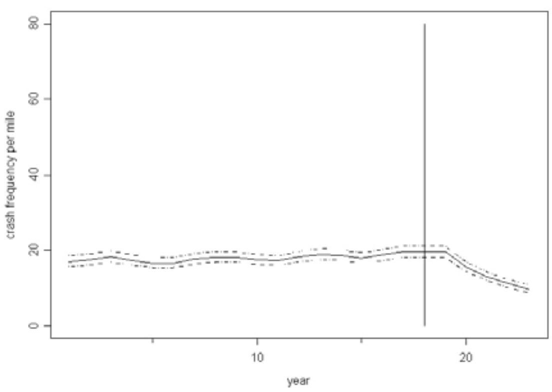
Expected yearly crash frequency per mile at Site 12



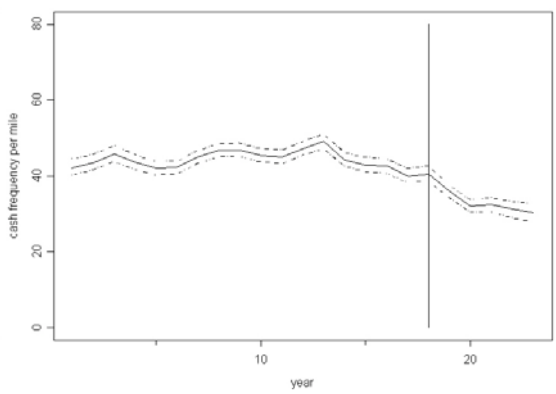
Expected yearly crash frequency per mile at Site 29



Expected yearly crash frequency per mile at Site 13



Expected yearly crash frequency per mile at Site 30



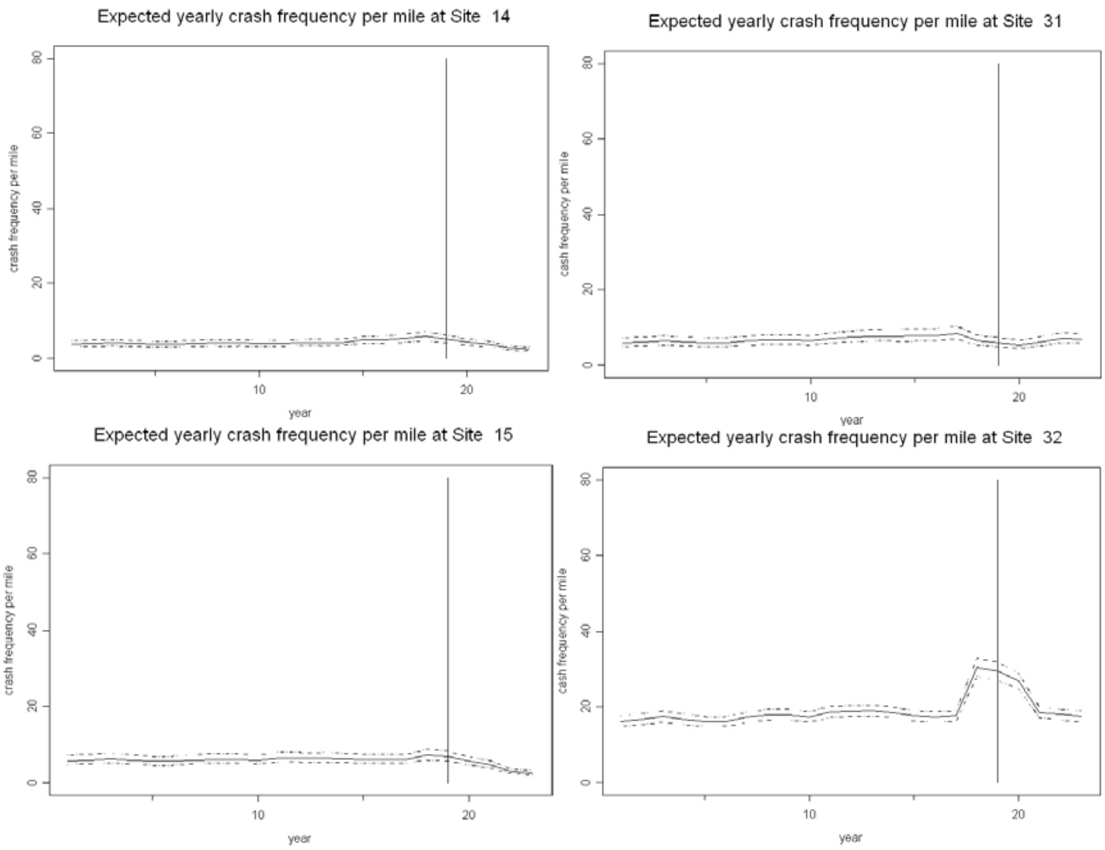


Figure 4: Posterior mean and 95% credible set of the expected crash frequency per year and mile for each site in the study. Years preceding the completion of the intervention are to the left of the vertical line in each plot.

From the figure, we see that there was an estimated reduction in crash frequency per mile at most sites. The reduction appears to be more pronounced at sites that underwent the conversion from four to three lanes. Notice that the 95% credible sets for expected crash frequencies are in general rather narrow. Thus, we are confident that site-specific expected annual crash frequency per mile is estimated with a good degree of confidence.

Table 9 shows, for each site in the study, the posterior mean and 95% credible set for each of the following quantities:

- The average expected crash frequency per year and mile, during years preceding the completion of the four to three-lane conversion. Formally, the average for the *i*th site is computed as:

$$\frac{\sum_{j=Start.year}^{Comp.year} \mu_{ij} / L_{ij}}{n_i},$$

  where L_i is the length of the *i*th site at the year j, and n_i is the number of years of traffic safety information for the site during the years preceding the conversion. Here, μ_{ij} is the expected number of crashes per mile in *j*th year for the *i*th site and is obtained directly from the Poisson model fitted earlier.
- The average expected crash frequency per year and mile, during years following the completion of the four to three-lane conversion. Formally, the average for the

*i*th site is computed as above, but now the average is taken over the years that follow the completion of the intervention.

- The difference in expected crash frequency in the two periods.
- The reduction, expressed as a percentage, in the expected annual crash frequency per mile at each of the sites between the "After" and the "Before" periods.

Site	Expected annual crash frequency per mile		Difference in crash frequency	Percent reduction in crash frequency
	Before	After		
1	43 (40.9,46.5)	21 (20.2,23.3)	-22.1 (-24.5,-19.7)	50.4 (47.4,53.6)
2	21 (20.2,23.1)	19 (17.8,21.0)	-2.2 (-3.1,-1.3)	10.2 (6.2,14.3)
3	5 (4.9,6.6)	3 (2.6,3.6)	-2.6 (-3.1,-2.2)	46.0 (43.7,48.3)
4	22 (21.3,24.1)	15 (14.7,17.2)	-6.9 (-7.7,-6.1)	30.4 (27.4,33.4)
5	25 (22.1,29.0)	20 (18.0,24.0)	-4.6 (-5.8,-3.6)	18.1 (14.6,21.5)
6	21 (19.3,23.8)	19 (17.8,21.8)	-2.0 (-2.9,-1.2)	9.2 (5.3,13.0)
7	9 (7.3,12.6)	5 (3.8,6.7)	-4.5 (-5.9,-3.4)	47.4 (45.2,49.8)
8	25 (22.9,27.5)	13 (11.9,14.4)	-11.9 (-13.4,-10.9)	47.5 (45.3,50.0)
9	53 (50.2,55.9)	35 (33.1,37.5)	-17.7 (-19.6,-15.9)	33.3 (30.8,36.0)
10	28 (24.4,34.1)	20 (17.5,24.0)	-8.0 (-9.7,-6.5)	28.2 (25.3,31.1)
11	17 (15.0,19.3)	7 (6.8,8.6)	-9.4 (-10.7,-8.2)	55.0 (53.4,56.5)
12	35 (33.6,37.6)	24 (22.6,25.6)	-11.6 (-13.0,-10.3)	32.6 (30.0,35.5)
13	18 (16.8,19.4)	14 (13.6,15.8)	-3.4 (-4.1,-2.8)	18.8 (15.7,22.0)
14	4 (3.4,5.2)	3 (3.0,4.6)	-0.5 (-0.7,-0.4)	12.5 (9.2,15.9)
15	6 (5.1,7.6)	4 (3.9,5.7)	-1.5 (-1.9,-1.2)	24.3 (21.5,27.2)
18	24 (22.3,27.3)	23 (21.1,25.8)	-1.2 (-3.2,0.9)	4.7 (-3.9,12.4)
19	24 (23.6,26.0)	26 (24.6,27.7)	1.3 (0.3,2.1)	-5.1 (-8.6,-1.0)
20	1 (1.1,1.8)	2 (1.9,3.0)	1.0 (0.8,1.2)	-67.8 (-72.7,-62.8)
21	7 (6.7,8.1)	7 (6.5,7.9)	-0.2 (-0.5,0.1)	2.3 (-0.7,6.1)
22	14 (12.6,5.4)	13 (12.5,15.4)	-0.1 (-0.6,0.4)	0.4 (-2.7,4.4)
23	5 (4.4,6.6)	5 (4.6,7.0)	0.3 (0.0,0.5)	-5.1 (-8.7,-0.8)
24	6 (5.7,7.6)	5 (4.5,6.2)	-1.3 (-1.6,-1.0)	19.0 (15.4,22.7)
25	14 (13.4,16.3)	13 (12.2,15.0)	-1.2 (-1.8,-0.5)	7.8 (3.6,12.1)
26	34 (32.4, 36.7)	32 (30.6,34.8)	-1.9 (-3.4,-0.5)	5.4 (1.5,9.6)
27	5 (4.6,6.7)	5 (4.4,6.4)	-0.3 (-0.5,-0.1)	4.9 (1.2,9.0)
28	9 (8.0,10.8)	8 (7.2,9.8)	-0.8 (-1.1,-0.5)	8.7 (5.0,12.5)
29	20 (19.0,21.6)	16 (15.4,18.0)	-3.5 (-4.3,-2.8)	17.5 (14.2,21.1)
30	44 (42.7,45.8)	33 (31.6,35.2)	-10.6 (-12.1,-9.1)	24.1 (20.9,27.4)
31	6 (5.6,8.3)	6 (5.3,7.9)	-0.4 (-0.7,-0.1)	5.4 (1.7,9.5)
32	18 (17.3,20.1)	22 (20.8,24.2)	3.9 (3.0,4.7)	-20.7 (-24.6,-16.0)

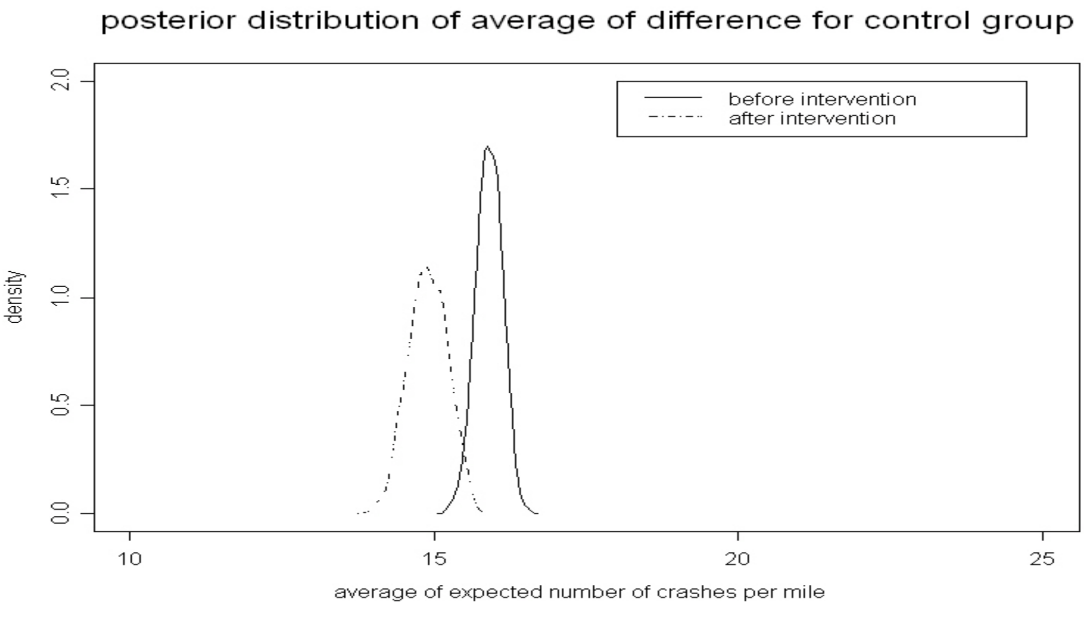
Table 9: Expected annual crash frequency per mile per site during "Before" and "After" periods, difference in expected crash frequency and percent reduction in crash frequency.

Table 10 shows the overall effect of the intervention. We computed the posterior distribution of expected annual crash frequencies per mile for all treated and all control sites over the years preceding and following the intervention. The four posterior distributions are shown in Figure 5.

	Crash frequency per mile before completion	Crash frequency per mile after completion
All control sites	16 (15.5, 16.3)	15 (14.3, 15.5)
All treatment sites	23 (21.9, 23.2)	15 (14.6, 15.9)

Table 10: Posterior mean and 95% credible set of the average (over treated and control sites) of the expected annual crash frequency per mile during the years preceding and following the four to three-lane conversion.

From the figures, note that while the expected annual crash frequency per mile has decreased at all sites, the reduction is significantly more pronounced at sites that underwent the conversion. The posterior distributions shown in Figure 5 are narrow, indicating that the posterior mean is a reliable summary of the distribution of likely values of expected crash frequencies.



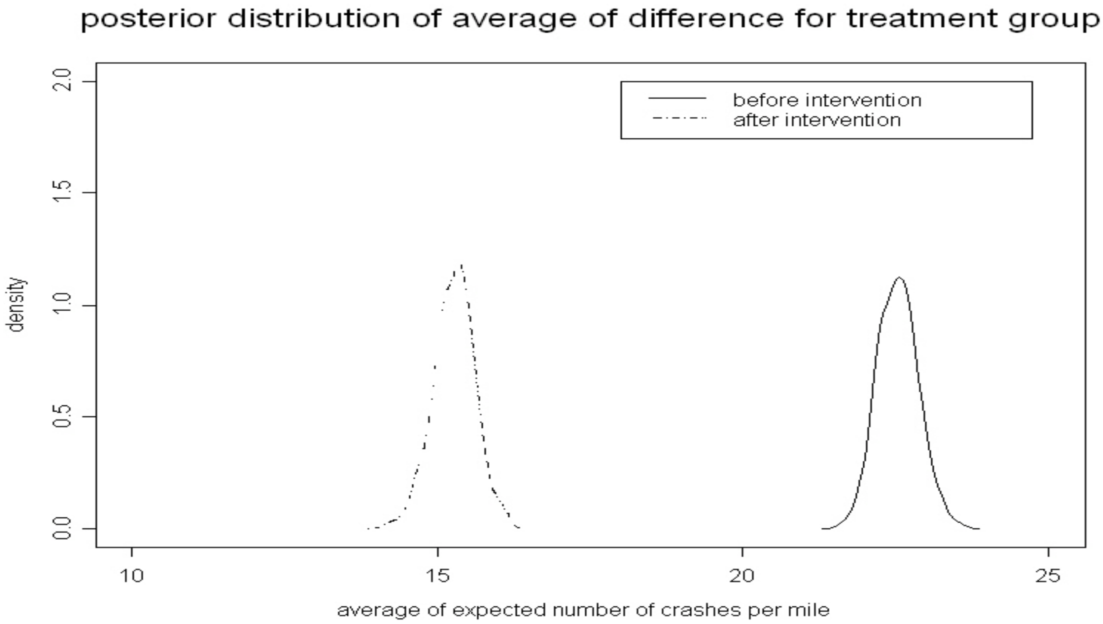


Figure 5: Posterior distributions of the average (across treated and control sites) expected annual crash frequencies per mile during the years preceding and following the completion of the intervention.

Because the traffic volume at the different sites varied across sites both during the “Before” and “After” periods and also because volume appears to be on the rise at all sites, we recomputed the posterior distributions of the expected annual crash frequencies per site and mile during each of the two periods, but now re-normalizing each site to a common 100,000,000 AADT, which actually is the expected annual crash rate per mile. The results of those calculations are shown in Tables 11 (averages over all sites in each of the two groups) and 12 (site-specific values). Notice that the expected crash frequencies as well as the percent reductions in the expected crash frequencies differ from those presented in Tables 9 and 10. This is to be expected given the variability and evolution in traffic volumes.

	Expected annual crash rate per mile before completion	Expected annual crash rate per mile after completion
All control sites	652 (617,683)	486 (466,506)
All treatment sites	792 (758,826)	442 (422,462)

Table 11: Posterior mean and 95% credible set of the average (over treated and control sites) of the expected annual crash rate per mile during the years preceding and following the four to three-lane conversion.

Site	Expected annual crash rate per mile		Difference in crash rate	Percent reduction in crash rate
	Before	After		
1	2068(1873,2263)	810(751,868)	-1258(-1439,-1078)	60.8(57.3,64.2)
2	631(590,680)	444(408,482)	-187(-217,-158)	29.6(25.7,33.1)
3	180(136,223)	117(100,137)	-63(-100,-23)	33.8(15.8,46.4)
4	807(721,940)	531(496,580)	-276(-391,-204)	33.9(27.2,42.5)
5	776(616,939)	510(441,587)	-265(-397,-115)	33.7(18.8,43.3)
6	708(634,782)	499(453,555)	-210(-235,-181)	29.6(26.6,32.6)
7	551(421,726)	278(212,369)	-274(-353,-208)	49.6(47.5,51.8)
8	1125(1028,1233)	584(530,638)	-541(-608,-494)	48.1(45.8,50.4)
9	1256(1194,1334)	707(660,750)	-549(-605,-496)	43.7(41.0,46.6)
10	866(745,1040)	511(439,601)	-354(-426,-299)	40.9(38.5,43.3)
11	682 (602,771)	383(337,434)	-299(-342,-257)	43.9(41.8,46.0)
12	1231(1162,1298)	681(641,724)	-551(-594,-501)	44.7(42.6,46.8)
13	564 (523,607)	315(292,340)	-249(-274,-223)	44.1(41.5,46.7)
14	190(149,228)	115(90,138)	-75(-91,-59)	39.5(37.2,41.8)
15	239(195,290)	147(121,178)	-92(-114,-73)	38.5(36.0,40.9)
18	1317(1141,1483)	864(781,950)	-453(-584,-302)	34.2(26.0,40.9)
19	667(632,707)	538(508,573)	-130(-154,-107)	19.4(16.5,23.0)
20	370(294,468)	236(192,298)	-134(-183,-93)	36.0(31.3,40.8)
21	437(318,567)	348(314,384)	-89(-209,12)	17.6(-3.2,37.5)
22	682(525,888)	552(494,611)	-130(-304,11)	17.1(-2.2,35.0)
23	245(201,301)	200(162,250)	-46(-56,-36)	18.6(15.8,21.8)
24	528(452,609)	384(324,447)	-143(-175,-119)	27.2(23.8,30.5)
25	748(675,821)	547(489,598)	-201(-236,-170)	26.9(23.5,30.2)
26	867(811,927)	649(607,691)	-218(-254,-180)	25.1(21.7,28.9)
27	170(137,199)	143(114,168)	-27(-36,-20)	15.9(12.7,19.5)
28	1026(883,1191)	754(629,862)	-273(-335,-214)	26.5(22.7,30.4)
29	822(771,885)	624(579,675)	-197(-225,-168)	24.0(20.7,27.3)
30	1047(1009,1085)	795(750,835)	-253(-290,-217)	24.1(21.0,27.4)
31	243(200,295)	190(156,231)	-54(-69,-40)	22.1(19.4,25.3)
32	613(564,659)	474(437,507)	-140(-160,-121)	22.8(20.0,25.9)

Table 12: Expected annual crash frequency per mile per site for 1,000,000 AADT during "Before" and "After" periods, difference in expected crash frequency and percent reduction in crash frequency.

4.2 Forecasting expected crash numbers at the 30 study sites

In this section, we compute the posterior predictive distributions of:

- Expected annual crash frequency per mile at each site

- Difference in expected crash frequency per mile between treatment and control sites defined as

$$\frac{100 \times \mu_{it,[treatment]}}{\mu_{it,[control]}}$$

which we label “Relative expected crash frequencies” and denote D_1 .

- Difference in expected crash frequency per mile between treatment and control sites defined as

$$\frac{100 \times \left(\mu_{it,[control]} - \mu_{it,[treatment]}\right)}{\mu_{it,[control]}}$$

which we label “Percent reduction in expected crash frequency” and denote D_2.

Note that $D_2 = 1 - D_1$.

In the expressions above, μ_{it} is defined as

$$\mu_{it} = \left(\frac{\lambda_{it} v_{it}}{1000 L_{it}}\right).$$

That is, μ_{it} denotes the expected number of crashes (per 1,000 ADT) at the *i*th site during the *t*th time period, while L_{it} , the length of road segment at the *i*th site at *t*th time period, is defined similarly to the previous part. Predictions are carried out for each of the 30 study sites during six periods: the months of January and June one, three and five years after the date the intervention was completed. If the intervention is effective, we expect both D_1 and D_2 to decrease over time. Ideally, we hope that D_1 will be less than 100 and that t D_2 will be negative. Even if this is not the case, the reason may be differences in the actual level of crash frequency at each treatment site relative to its control site, so what we are most interested in is the *evolution* of the two metrics rather than their actual values.

Since site 5 and site 8 had the same matched control site but the completion date of intervention at the two sites was different, we used for prediction site 22 in both cases but changed the completion date at the control accordingly, to correspond to the intervention data at each of the treatment sites.

Because in some cases we predict crashes at a time that exceeds the study period (out of sample prediction), we have no ADT information to insert into the model to carry out the forecasting. In those cases, we use the value of MADT in the last January or June in the dataset at each site. Where available, we use the observed (or estimated) MADT at that time. For example, if the completion date is June, 1993, we calculate the predicted number of crashes per mile and differences in crash numbers at the site for January /June in 1995, January/June in 1997, and January/June in 1999. If the completion date is July, 2001, we calculate the predicted number of crashes and differences in crash numbers for January /June in 2003, and January/June in 2005, and January/June in 2007. In this case, we use the MADT in January/June in 2004 as the MADT in January/June in 2005 and 2007.

The tables that follow, Table 13 – Table 18 show, for each treated site and its matched pair, the predicted expected annual crash frequency per mile during each of the prediction periods. In each case, we provide the posterior mean as well as the 95% credible set.

From the tables, we see that the predicted expected crash frequency per mile at each of the sites is consistent with the numbers of crashes per mile observed at each of the sites, at least when the prediction period was included in the dataset. Hence we are confident that the model predictions are reasonable.

Road	Traffic Volume (MADT)	Mean (0.025, 0.975)	Road	Traffic Volume (MADT)	Mean (0.025, 0.975)
1	190000	3.10 (2.90,3.30)	18	168000	2.15 (1.93,2.38)
2	328000	1.46 (1.34,1.59)	19	363000	2.20 (2.04,2.37)
3	194000	0.26 (0.22,0.30)	20	76000	0.21 (0.17,0.26)
4	221000	1.33 (1.23,1.45)	21	154000	0.62 (0.55,0.68)
5	304000	1.76 (1.52,2.04)	22	188000	1.20 (1.06,1.34)
6	290000	1.55 (1.40,1.71)	23	212000	0.48 (0.39,0.58)
7	130000	0.54 (0.40,0.70)	24	102000	0.50 (0.42,0.59)
8	175000	1.51 (1.37,1.64)	25	191000	1.36 (1.22,1.50)
9	381000	3.60 (3.37,3.85)	26	389000	3.10 (2.90,3.31)
10	298000	1.87 (1.57,2.20)	27	298000	0.53 (0.43,0.63)
11	62000	0.32 (0.28,0.37)	28	82000	0.76 (0.65,0.89)
12	242000	2.18 (2.03,2.35)	29	196000	1.52 (1.41,1.63)
13	361000	1.54 (1.41,1.68)	30	302000	2.99 (2.82,3.17)
14	266000	0.38 (0.30,0.47)	31	257000	0.58 (0.46,0.72)
15	267000	0.48 (0.39,0.57)	32	249000	1.53 (1.41,1.67)

Table13: Posterior predictive distribution of the expectation of the number of crashes per mile at each study site for the month of January, one year after completion of the intervention

Road	Traffic Volume (MADT)	Mean (0.025, 0.975)	Road	Traffic Volume (MADT)	Mean (0.025, 0.975)
1	222000	2.56 (2.40,2.74)	18	196000	1.86 (1.68,2.06)
2	382000	1.21 (1.11,1.32)	19	423000	1.91 (1.77,2.06)
3	226000	0.22 (0.19,0.25)	20	83000	0.17 (0.13,0.21)
4	257000	1.10 (1.01,1.20)	21	179000	0.54 (0.48,0.59)
5	354000	1.46 (1.25,1.70)	22	220000	1.04 (0.92,1.17)
6	338000	1.28 (1.16,1.42)	23	247000	0.42 (0.34,0.50)
7	152000	0.45 (0.33,0.59)	24	119000	0.44 (0.36,0.51)
8	204000	1.25 (1.14,1.37)	25	223000	1.18 (1.06,1.31)
9	416000	2.79 (2.60,2.99)	26	424000	2.52 (2.35,2.69)
10	348000	1.55 (1.30,1.82)	27	326000	0.43 (0.35,0.52)
11	68000	0.25 (0.22,0.28)	28	95000	0.66 (0.56,0.77)
12	282000	1.81 (1.69,1.94)	29	229000	1.32 (1.22,1.41)
13	420000	1.28 (1.17,1.39)	30	352000	2.60 (2.45,2.76)
14	310000	0.31 (0.25,0.39)	31	299000	0.50 (0.40,0.62)
15	292000	0.37 (0.30,0.44)	32	272000	1.25 (1.14,1.35)

Table 14: Posterior predictive distribution of the expectation of the number of crashes at each study site for the month of June, per mile, one year after completion of the intervention

Road	Traffic Volume (MADT)	Mean (0.025, 0.975)	Road	Traffic Volume (MADT)	Mean (0.025, 0.975)
1	203000	2.31 (2.14,2.49)	18	179000	2.02 (1.82,2.22)
2	328000	1.03 (0.93,1.13)	19	363000	1.94 (1.77,2.12)
3	196000	0.18 (0.16,0.22)	20	77000	0.18 (0.15,0.23)
4	223000	0.94 (0.85,1.04)	21	155000	0.55 (0.49,0.61)
5	307000	1.25 (1.06,1.45)	22	190000	1.07 (0.94,1.21)
6	290000	1.09 (0.97,1.21)	23	212000	0.42 (0.35,0.51)
7	122000	0.27 (0.20,0.36)	24	101000	0.44 (0.37,0.52)
8	174000	1.06 (0.95,1.16)	25	188000	1.18 (1.05,1.32)
9	381000	2.53 (2.32,2.74)	26	389000	2.73 (2.52,2.96)
10	298000	1.26 (1.05,1.50)	27	281000	0.44 (0.35,0.53)
11	64000	0.23 (0.20,0.27)	28	81000	0.67 (0.57,0.78)
12	320000	2.03 (1.86,2.21)	29	196000	1.34 (1.23,1.45)
13	358000	1.07 (0.97,1.18)	30	322000	2.82 (2.62,3.02)
14	224000	0.22 (0.18,0.28)	31	310000	0.62 (0.49,0.76)
15	195000	0.25 (0.21,0.30)	32	249000	1.35 (1.22,1.50)

Table 15: Posterior predictive distribution of the expectation of the number of crashes at each study site for the month of January, per mile, three years after completion of the intervention

Road	Traffic Volume (MADT)	Mean (0.025, 0.975)	Road	Traffic Volume (MADT)	Mean (0.025, 0.975)
1	236000	1.91 (1.77,2.07)	18	208000	1.75 (1.58,1.95)
2	382000	0.85 (0.77,0.94)	19	423000	1.68 (1.54,1.85)
3	229000	0.15 (0.13,0.18)	20	84000	0.15 (0.12,0.19)
4	260000	0.78 (0.70,0.86)	21	181000	0.48 (0.42,0.54)
5	357000	1.03 (0.88,1.21)	22	222000	0.93 (0.81,1.05)
6	338000	0.90 (0.80,1.00)	23	247000	0.37 (0.30,0.44)
7	142000	0.23 (0.17,0.30)	24	118000	0.38 (0.32,0.45)
8	203000	0.88 (0.79,0.97)	25	220000	1.03 (0.91,1.15)
9	416000	1.96 (1.79,2.13)	26	424000	2.22 (2.04,2.41)
10	348000	1.04 (0.87,1.25)	27	307000	0.36 (0.29,0.43)
11	70000	0.18 (0.16,0.21)	28	94000	0.58 (0.49,0.68)
12	374000	1.68 (1.53,1.84)	29	229000	1.16 (1.06,1.26)
13	417000	0.88 (0.80,0.98)	30	375000	2.45 (2.27,2.64)
14	261000	0.19 (0.15,0.23)	31	362000	0.54 (0.43,0.66)
15	213000	0.20 (0.16,0.23)	32	272000	1.10 (0.99,1.21)

Table 16: Posterior predictive distribution of the expectation of the number of crashes at each study site for the month of June, per mile, three years after completion of the intervention

Road	Traffic Volume (MADT)	Mean (0.025, 0.975)	Road	Traffic Volume (MADT)	Mean (0.025, 0.975)
1	215000	1.72 (1.56,1.90)	18	253000	2.52 (2.26,2.80)
2	328000	0.72 (0.64,0.81)	19	363000	1.71 (1.53,1.91)
3	196000	0.13 (0.11,0.15)	20	77000	0.16 (0.13,0.20)
4	223000	0.66 (0.58,0.75)	21	155000	0.49 (0.43,0.55)
5	307000	0.88 (0.74,1.03)	22	190000	0.94 (0.81,1.08)
6	290000	0.76 (0.67,0.86)	23	212000	0.37 (0.30,0.45)
7	152000	0.31 (0.23,0.41)	24	101000	0.39 (0.32,0.46)
8	176000	0.72 (0.63,0.81)	25	190000	1.05 (0.93,1.20)
9	385000	1.79 (1.60,1.99)	26	392000	2.44 (2.20,2.70)
10	298000	0.88 (0.73,1.06)	27	281000	0.39 (0.31,0.47)
11	65000	0.16 (0.14,0.19)	28	82000	0.60 (0.50,0.71)
12	324000	1.44 (1.29,1.61)	29	198000	1.19 (1.07,1.32)
13	362000	0.76 (0.67,0.85)	30	325000	2.51 (2.27,2.77)
14	224000	0.16 (0.12,0.20)	31	310000	0.55 (0.43,0.68)
15	195000	0.18 (0.14,0.22)	32	427000	2.05 (1.80,2.31)

Table 17: Posterior predictive distribution of the expectation of the number of crashes at each study site for the month of January, per mile, five years after completion of the intervention

Road	Traffic Volume (MADT)	Mean (0.025, 0.975)	Road	Traffic Volume (MADT)	Mean (0.025, 0.975)
1	251000	1.43 (1.28,1.58)	18	294000	2.19 (1.95,2.44)
2	382000	0.60 (0.52,0.67)	19	423000	1.49 (1.32,1.67)
3	229000	0.11 (0.09,0.13)	20	84000	0.13 (0.10,0.17)
4	260000	0.55 (0.48,0.62)	21	181000	0.42 (0.37,0.48)
5	357000	0.73 (0.60,0.86)	22	222000	0.82 (0.70,0.95)
6	338000	0.63 (0.55,0.72)	23	247000	0.33 (0.26,0.40)
7	177000	0.26 (0.18,0.34)	24	118000	0.34 (0.28,0.40)
8	205000	0.59 (0.52,0.67)	25	222000	0.91 (0.80,1.05)
9	420000	1.39 (1.23,1.55)	26	428000	1.98 (1.78,2.20)
10	348000	0.73 (0.60,0.88)	27	307000	0.32 (0.25,0.39)
11	71000	0.13 (0.11,0.15)	28	95000	0.52 (0.43,0.62)
12	377000	1.19 (1.05,1.34)	29	231000	1.04 (0.92,1.15)
13	422000	0.63 (0.55,0.71)	30	379000	2.18 (1.97,2.42)
14	261000	0.13 (0.10,0.16)	31	362000	0.48 (0.37,0.59)
15	213000	0.14 (0.11,0.17)	32	466000	1.67 (1.46,1.88)

Table 18: Posterior predictive distribution of the expectation of the number of crashes at each study site for the month of June, per mile, five years after completion of the intervention

From the results shown in the tables above, we see that in general, as time goes by, the expected number of crashes per mile at each site in the treatment group continues to decrease faster than the number at the corresponding paired site in the control group.

It is also of interest to predict the difference in expected crash numbers between treatment and control sites that can be expected after the intervention is completed. We obtained the posterior predictive distributions of the differences in crash numbers per mile that were defined earlier and that we denoted D_1 and D_2. Both definitions of the expected difference attempt to compare what can be expected in terms of crash numbers per mile at treatment sites when compared to control sites.

Results are presented in Tables 19 and 20. In the tables, each column represents a prediction period. Entries in the table are the means of the posterior predictive distributions of the different metrics and in parenthesis, the 95% credible sets for the metrics. For example, if the 95% credible set is (47, 62) we say that with 95% probability, the predicted number of crashes at the treatment site is between 47% and 62% of the predicted number of crashes at the matched control site.

Relative expected crash frequencies (D_1) or expected frequency at converted site relative to expected frequency at corresponding paired site. Entries are: Mean (0.025, 0.975)						
Paired Road	Jan., 1 year	June, 1 year	Jan., 3 year	June, 3 year	Jan., 5 year	June, 5 year
1 and 18	144.6 (129.0,162.0)	137.8 (122.7,154.4)	114.9 (101.8,129.2)	109.5 (96.4,123.5)	68.7 (59.3,79.0)	65.5 (56.1,75.6)
2 and 19	66.8 (59.9,74.4)	63.6 (56.9,71.0)	53.1 (46.4,60.1)	50.6 (44.1,57.6)	42.2 (35.7,49.5)	40.2 (33.7,47.6)
3 and 20	127.6 (96.5,162.8)	129.9 (98.2,165.7)	101.4 (75.8,129.2)	103.2 (76.9,131.8)	80.6 (59.3,104.5)	82.1 (59.9,106.7)
4 and 21	216.6 (191.5,245.6)	206.4 (181.7,234.3)	172.1 (148.8,197.7)	164.1 (140.7,189.8)	136.9 (113.8,162.4)	130.5 (107.4,155.7)
5 and 22	147.7 (122.4,177.8)	140.8 (116.6,169.7)	117.4 (96.2,143.4)	111.9 (90.9,137.3)	93.4 (74.6,116.5)	89.0 (70.4,111.5)
6 and 23	325.6 (263.0,399.7)	310.4 (250.4,380.8)	258.7 (206.8,319.2)	246.6 (196.0,305.1)	205.7 (161.2,256.2)	196.1 (153.2,244.6)
7 and 24	108.7 (77.0,147.7)	103.6 (73.3,141.2)	62.9 (44.0,85.7)	59.9 (41.8,82.0)	80.7 (55.5,111.6)	77.0 (52.8,106.9)
8 and 25	111.5 (97.1,126.8)	106.2 (92.4,121.1)	89.8 (77.2,103.3)	85.6 (73.1,99.1)	68.3 (56.5,81.1)	65.1 (53.5,77.9)
9 and 26	116.5 (106.8,127.0)	111.0 (101.6,121.4)	92.5 (82.7,102.9)	88.2 (78.2,98.8)	73.6 (63.2,84.7)	70.2 (59.5,81.6)
10 and 27	358.6 (274.8,459.8)	365.1 (279.6,468.5)	290.1 (221.4,378.1)	295.3 (223.8,386.2)	230.7 (172.0,306.0)	234.9 (175.1,312.7)
11 and 28	42.5 (34.7,52.3)	38.0 (30.8,46.5)	35.0 (27.8,43.1)	31.2 (24.8,38.7)	27.8 (21.7,35.0)	24.8 (19.2,31.4)
12 and 29	144.2 (131.0,159.1)	137.4 (124.3,152.3)	151.8 (134.9,170.3)	144.8 (127.9,163.6)	120.8 (103.1,140.1)	115.1 (97.3,134.5)
13 and 30	51.5 (46.3,56.8)	49.1 (44.1,54.2)	37.9 (33.6,42.7)	36.1 (31.8,40.9)	30.1 (25.7,35.1)	28.7 (24.2,33.7)
14 and 31	66.1 (47.7,88.9)	63.0 (45.4,84.7)	36.7 (26.5,49.6)	34.9 (25.2,47.3)	29.1 (20.8,40.2)	27.8 (19.7,38.4)
15 and 32	31.3 (25.5,38.0)	29.8 (24.2,36.3)	18.6 (14.9,23.0)	17.8 (14.1,21.9)	8.6 (6.7,10.8)	8.2 (6.4,10.4)

Table 13: Posterior predictive distribution for D_1

What do we expect to see in the entries of the table? We do not focus on the values of the metrics themselves since those mean little without a reference to the actual crash frequency level at each site. Instead, we focus on the *trend* of D_1 and D_2 over time, and note the following:

- As anticipated, D_1 decreases as time goes by. This is consistent with the steeper negative slope of log crash rate on time after completion of the conversion at converted sites.
- Also as anticipated, D_2 increases as time goes by. This is due to the same reason.

Percent reduction in crash frequency at converted relative to control sites (D_2). Entries are: Mean (0.025, 0.975)						
Paired Road	Jan., 1 year	June, 1 year	Jan., 3 year	June, 3 year	Jan., 5 year	June, 5 year
1 and 18	-44.6 (-62.0,-29.0)	-37.8 (-54.4,-22.7)	-14.9 (-29.2,-1.8)	-9.5 (-23.5,3.6)	31.3 (21.0,40.7)	34.5 (24.4,43.9)
2 and 19	33.2 (25.6,40.1)	36.4 (29.0,43.1)	46.9 (39.9,53.6)	49.4 (42.4,55.9)	57.8 (50.5,64.3)	59.8 (52.4,66.3)
3 and 20	-27.6 (-62.8,3.5)	-29.9 (-65.7,1.8)	-1.4 (-29.2,24.2)	-3.2 (-31.8,23.1)	19.4 (-4.5,40.7)	17.9 (-6.7,40.1)
4 and 21	-116.6 (-145.6,-91.5)	-106.4 (-134.3,-81.7)	-72.1 (-97.7,-48.8)	-64.1 (-89.8,-40.7)	-36.9 (-62.4,-13.8)	-30.5 (-55.7,-7.4)
5 and 22	-47.7 (-77.8,-22.4)	-40.8 (-69.7,-16.6)	-17.4 (-43.4,3.8)	-11.9 (-37.3,9.1)	6.6 (-16.5,25.4)	11.0 (-11.5,29.6)
6 and 23	-225.6 (-299.7,-163.0)	-210.4 (-280.8,-150.4)	-158.7 (-219.2,-106.8)	-146.6 (-205.1,-96.0)	-105.7 (-156.2,-61.2)	-96.1 (-144.6,-53.2)
7 and 24	-8.7 (-47.7,23.0)	-3.6 (-41.2,26.7)	37.1 (14.3,56.0)	40.1 (18.0,58.2)	19.3 (-11.6,44.5)	23.0 (-6.9,47.2)
8 and 25	-11.5 (-26.8,2.9)	-6.2 (-21.1,7.6)	10.2 (-3.3,22.8)	14.4 (0.9,26.9)	31.7 (18.9,43.5)	34.9 (22.1,46.5)
9 and 26	-16.5 (-27.0,-6.8)	-11.0 (-21.4,-1.6)	7.5 (-2.9,17.3)	11.8 (1.2,21.8)	26.4 (15.3,36.8)	29.8 (18.4,40.5)
10 and 27	-258.6 (-359.8,-174.8)	-265.1 (-368.5,-179.6)	-190.1 (-278.1,-121.4)	-195.3 (-286.2,-123.8)	-130.7 (-206.0,-72.0)	-134.9 (-212.7,-75.1)
11 and 28	57.5 (47.7,65.3)	62.0 (53.5,69.2)	65.0 (56.9,72.2)	68.8 (61.3,75.2)	72.2 (65.0,78.3)	75.2 (68.6,80.8)
12 and 29	-44.2 (-59.1,-31.0)	-37.4 (-52.3,-24.3)	-51.8 (-70.3,-34.9)	-44.8 (-63.6,-27.9)	-20.8 (-40.1,-3.1)	-15.1 (-34.5,2.7)
13 and 30	48.5 (43.2,53.7)	50.9 (45.8,55.9)	62.1 (57.3,66.4)	63.9 (59.1,68.2)	69.9 (64.9,74.3)	71.3 (66.3,75.8)
14 and 31	33.9 (11.1,52.3)	37.0 (15.3,54.6)	63.3 (50.4,73.5)	65.1 (52.7,74.8)	70.9 (59.8,79.2)	72.2 (61.6,80.3)
15 and 32	68.7 (62.0,74.5)	70.2 (63.7,75.8)	81.4 (77.0,85.1)	82.2 (78.1,85.9)	91.4 (89.2,93.3)	91.8 (89.6,93.6)

Table 14: Posterior predictive distribution for D_2, D_2 is the percent reduction crash frequency per mile, when comparison a treated site to its youked paired site during each of the six periods

Chapter 5 Conclusions and discussion

We have analyzed monthly crash data for 30 sites in Iowa, collected over 23 years between 1982 and 2004. Half of the sites were transformed from four to three lanes sometime during the study period while the other half (chosen to match the treatment sites) remained unchanged. The main objective of the study was to assess the effectiveness of the four-to-three lane change in terms of the expected crash frequency per mile at the sites.

We adopted a Bayesian approach throughout. A classical analysis could have been conducted and would have resulted in similar point estimates for model parameters. However, a classical analyst would have encountered some difficulties in estimating the variances of parameter estimates in the nonlinear model and would have had to resort to asymptotic approximations.

We fitted a hierarchical Poisson regression model to the crash frequency observed at each site. The log monthly crash rate per mile at each site was then modeled using a piecewise linear regression model with a change-point. The independent variables (or explanatory variables) in the change-point regression included the effects of the four seasons of the year, treatment, time and interactions of treatment and time. To estimate the association between log monthly crash rates and the explanatory variables, we added a random effect to account for overdispersion and for autocorrelation among observations obtained at the same site. We used proper but non-informative priors for all parameters in the model, and carried out all calculations using Markov chain Monte Carlo methods implemented in WinBUGS.
Our model permits accounting for temporal variation in traffic volume (e.g., Hauer, 1997) and also for the effect of season on crash frequency.

Traffic safety analysts often adopt a negative binomial (NB) model to represent number of crashes (e.g., Hauer, 1997; Huang et al., 2002). Marginally, the hierarchical Poisson model we chose is almost equivalent to the NB model. It is easy to show that a Poisson-Gamma model, in which a Poisson distribution is chosen for crash frequency and the expected frequency is in turn modeled using a Gamma distribution leads to a NB model marginally for the crash frequency. This can be shown by integrating the expected frequency from the joint distribution of observed and expected frequency. The resulting marginal distribution has the same form as the NB density. In our case, we modeled the log of the expected rate as a normal random variable, therefore assuming that the expected rate is distributed as an exponential-normal random variable. Thus, the marginal distribution of crash frequency, while not exactly NB, has a density that approximates the NB density.

Results suggest that the four-to-three lane conversion can be effective in reducing the number of crashes at least at the type of sites included in the Iowa study. The model that we fitted is a linear function of the log monthly crash rates at a site with time and other covariates and therefore results in an exponential decay of the expected number of crashes over time at the site after accounting for the potential effect of factors such as season. While we did not conduct a standard test for goodness of fit of the model, the estimates of parameters and the corresponding credible sets, as well as the results from posterior prediction at the individual sites and also for average treated and control sites suggest that the model is reasonable and fits the data well.

We assessed the effectiveness of the four to three lane conversion by comparing the average expected annual crash frequency per mile during years preceding and following the conversion at the site level and also as an average over all sites in each of the two groups (road diets and comparison sites). We found that the expected crash frequency decreases for both road diets and control sites in the “After” period when compared to the “Before” period. The decrease, however, is significantly larger in sites that underwent conversion. These results hold as well when expected crash frequency is expressed in a 100,000,000 AADT basis (i.e., when computed on a per rate basis).

We computed predictions for the 30 individual sites in the dataset for six selected months following the conversion at the treated sites: January and June one, three and five years after completion of the four to three-lane change. As expected given the estimated posterior distributions for model parameters, the expected number of crashes per mile during those months decreases more rapidly at converted sites than at control sites. In fact, assuming that the model continues to hold at future times not included in the dataset, the expected crash frequency at some converted sites appears to decrease rapidly enough to become lower than the expected frequency at the paired site even though the reverse was true before the intervention was implemented.

Our results strongly suggest that traffic safety is significantly improved by converting four lane roads to three lanes, at least in the State of Iowa and on the type of roads considered in this study. This is in contrast to the results reported in Huang et al. (2002). They studied crash frequency at several sites in the States of California and Washington, and concluded that the difference in reduction of crash frequency between road diets (12 sites) and control sites (25 sites) was not statistically significant. They included all crashes occurring three years before and three years after the conversion and excluded all crashes occurring during what they termed a “three-month transition” period. Because reliable AADT data were not available for several of the sites, a NB binomial model was fitted to crash frequency at eight road diets and 14 comparison sites. The expected frequency was then modeled as a linear function of city, site length, traffic volume, time, and treatment. Their results showed that while a reduction in crash frequency occurred at all sites, the differences between the “before” and “after” frequencies were not statistically distinguishable between treatment and control sites.

The differences between our analysis and the analysis performed by Huang et al. (2002) are several and may explain the diverging results. First, even the descriptive analysis of the “raw” data suggests that the effect of conversion in Iowa roads was much more dramatic than in the roads considered in the Huang et al. (2002) study. See, for example, the descriptive statistics presented in Table 3 of this report. Second, Huang et al. (2002) fitted an ordinary linear regression model to the expected crash frequencies, meaning that a single slope for expected frequency on time was assumed for the entire study period. We extended the model and allowed for different slopes during the “before” and the “after” periods explicitly by including a change-point in the model and for the interaction of treatment and slope. Notice that as a result, our model allows for a slight increase in crash frequency during the months immediately preceding the conversion and also during those months immediately following the conversion. Finally, we included a longer time series of crash frequencies as we included 23 years of data on almost all sites in the study. By analyzing monthly data, we were also able to account for seasonal variability in crash frequency and traffic volume; while a “must” in Iowa, where seasonal variation in driving conditions is marked, this may not be as critical in a study conducted in the northwestern region of the country.

And, in our data set, Site 20 behaved abnormal, many zero-valued-crashes time points in at this site, and tend to have more crashes after the year 2000. So, to be safe, we checked the database, and found that the values before the year 1993 were imaged, because before that time, the site doesn’t exist. However, it is easy to rerun WinBUGS and to recalculate all the results, and we found that the main conclusions still stay unchanged.

It is always possible to improve a statistical model. In our case, we might have included additional covariates to help explain the variability in crash numbers across sites. In addition, we could have included interaction terms to account for a potentially different effect of treatment on log crashes during the different seasons of the year. One disadvantage of including additional parameters in the model is that results become more difficult to interpret and we lose “error degrees of freedom”, meaning that all parameters are estimated less precisely. In this study, few observations were available for the period after the intervention was completed, so adding parameters for which data to estimate the parameters of the second component of the piecewise linear model were required might not have been a good idea.

It might be possible to improve on the precision of estimates if more crash data for the period following the intervention become available. The dataset contained very abundant information to estimate the parameters of the model pertinent to the first component of the piecewise linear regression. The second component, however is not as precisely estimated due to the few years of data available for most sites for the period following the four-to-three lane conversion.

Bibliography

Cowles, M.K.(2004). Review of WinBUGS 1.4.(Statistical Computing Software Reviews)(Software Reviews)(Product/Service Evaluation). *Journal of the American Statistical Association*, Vol. 58, i4 pp330(7).

Cowles, M.K. & Carlin, B.P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A comparative review. *Journal of the American Statistical Association*, Vol. 91, pp 434.

Gelman, A., et al.(2004). *Bayesian Data Analysis*, 2nd Edition. Chapman & Hill, London, England.

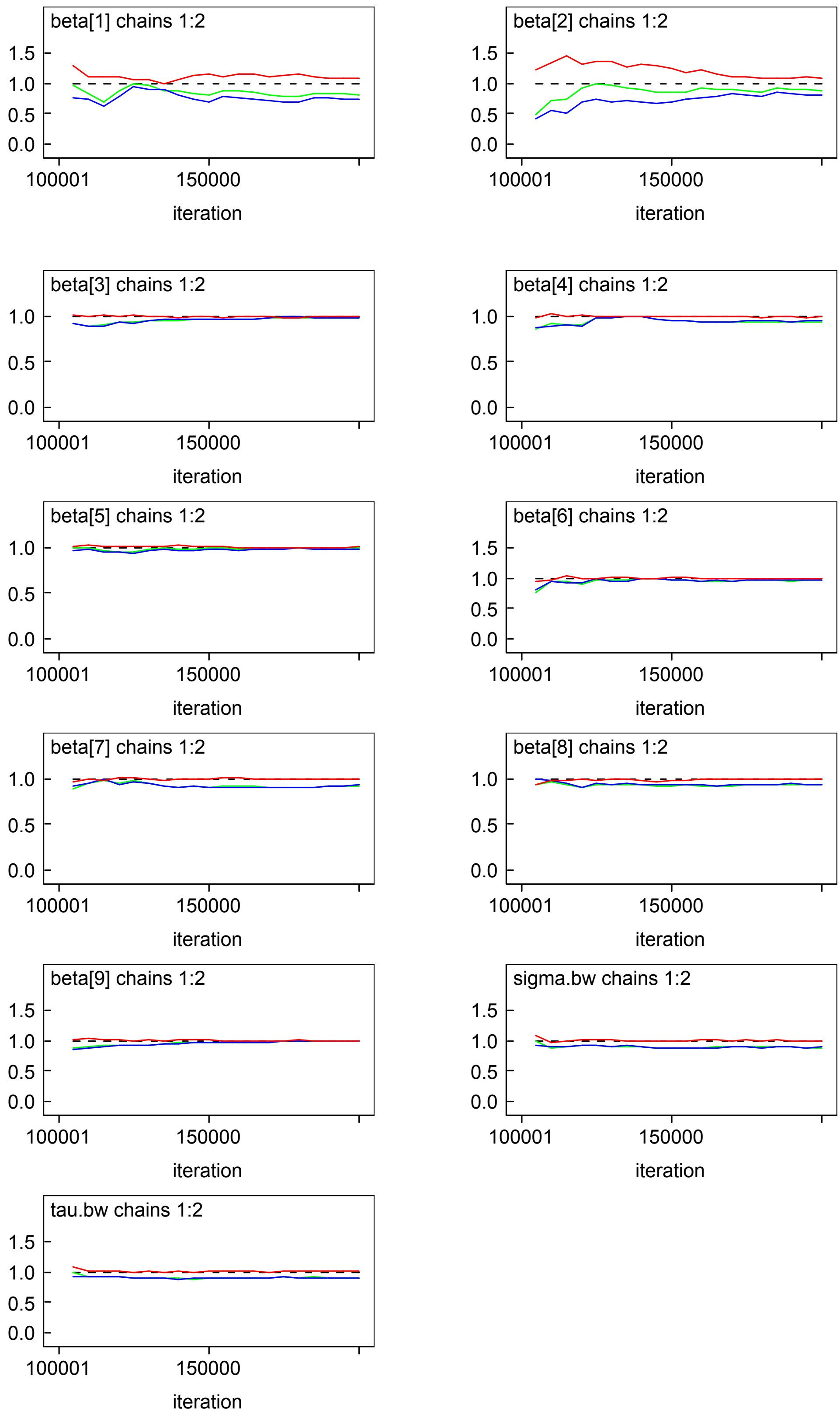
Hauer, E. (1997) *Observational Before-After Studies in Road Safety.* Pergamon, Oxford, U.K.

Huang, H.F., Stewart, J.R., and Zegeer, C.V. (2002) Evaluation of lane reduction "road diet" measures on crashes and injuries. *Transportation Research Record*, 1784, Paper number 02-2955: 80-90

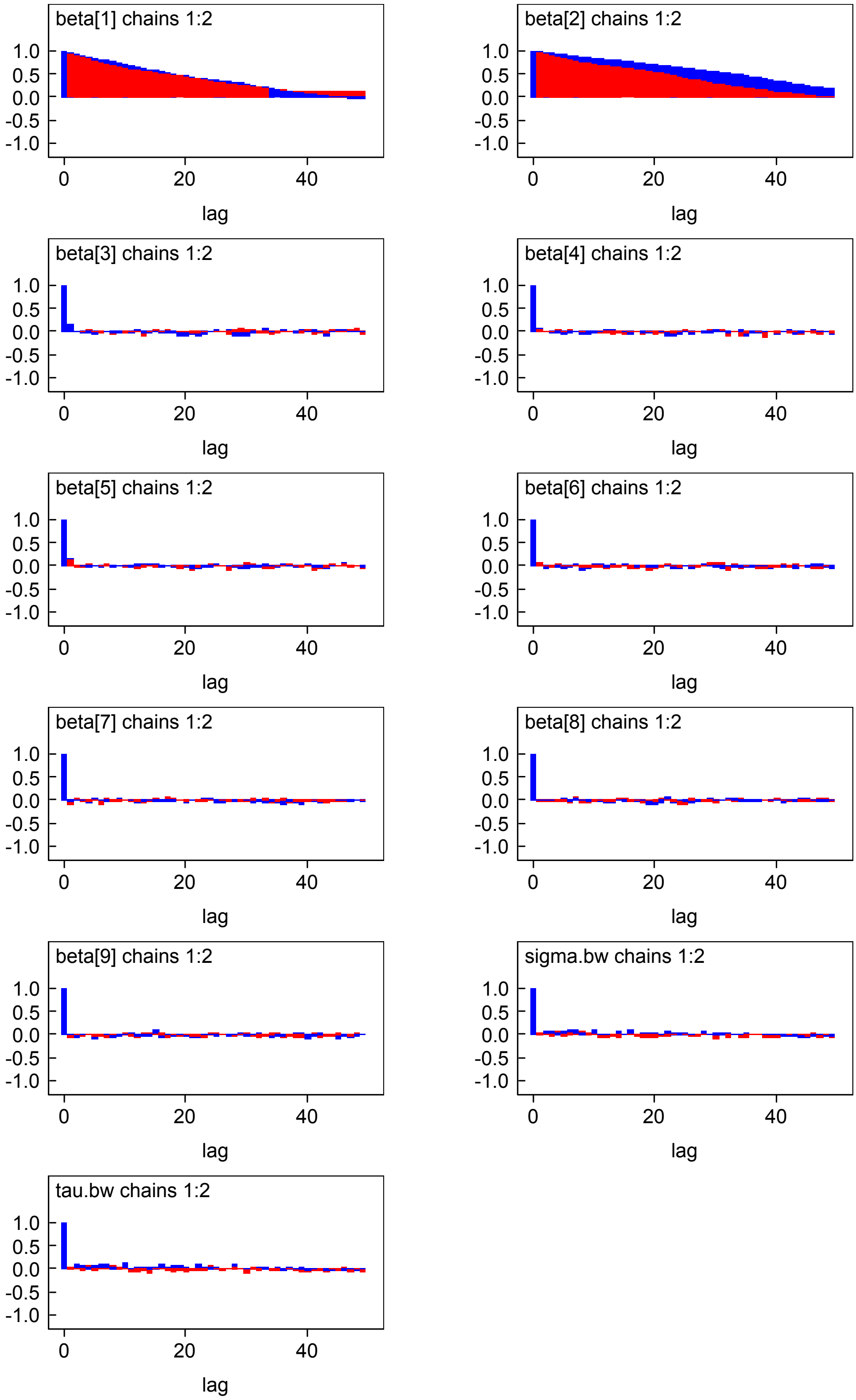
Pawlovich, M.D.(2003). *Evaluating Traffic Safety Network Screening: An initial framework utilizing the hierarchical Bayesian philosophy*, unpublished doctoral dissertation. Department of Civil and Construction Engineering, Iowa State University, Ames, Iowa.

Simonoff, J. S.(1996). *Smoothing Methods in Statistics*. Springer, New York, U.S.A..

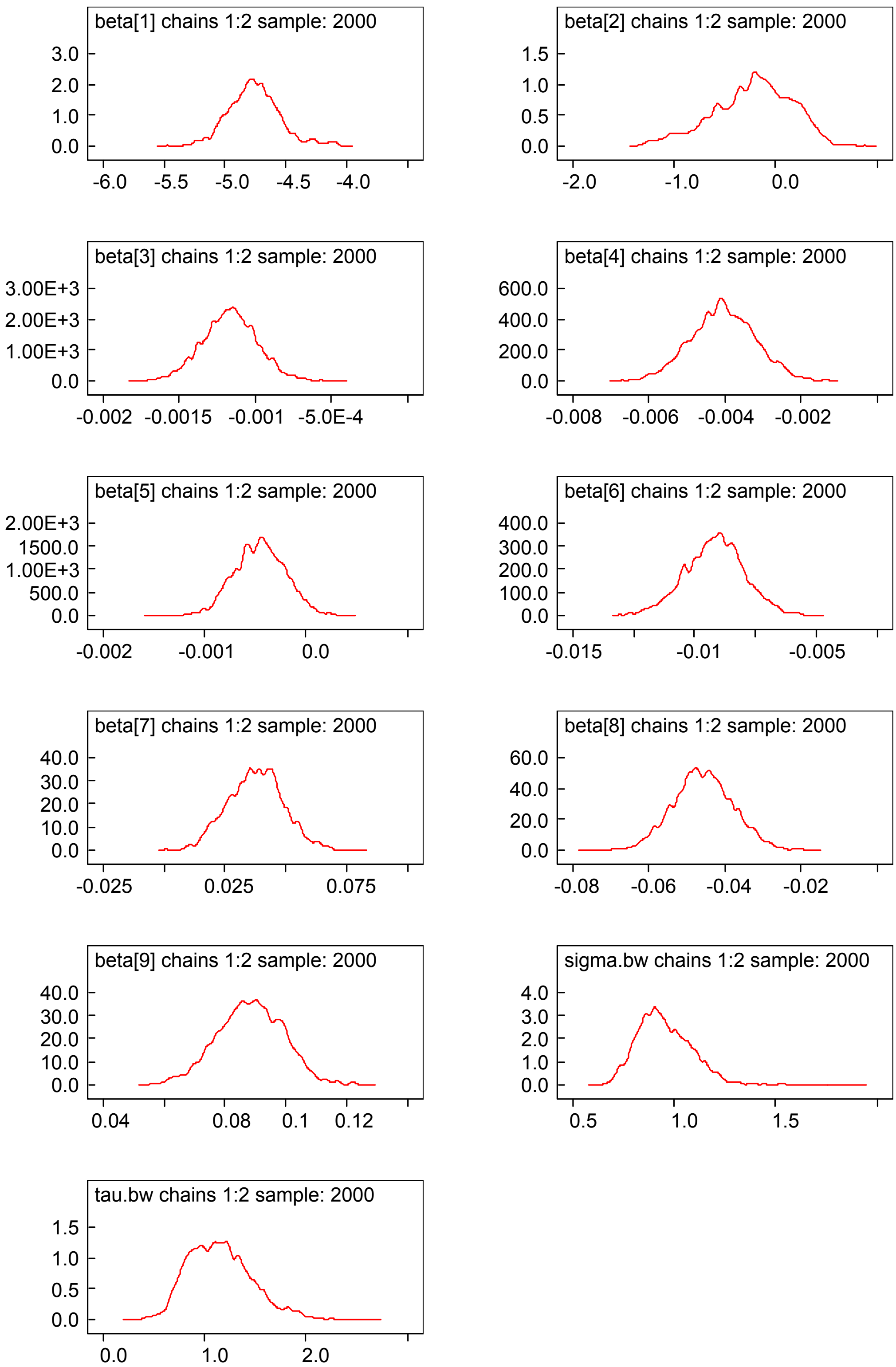
Appendix I: Gelman-Rubin Statistic Graphs for parameters



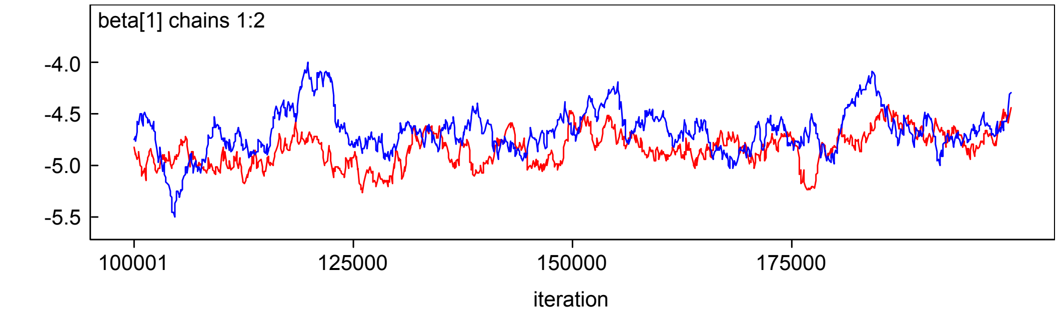
Appendix II: Autocorr. Statistic Graphs for parameters

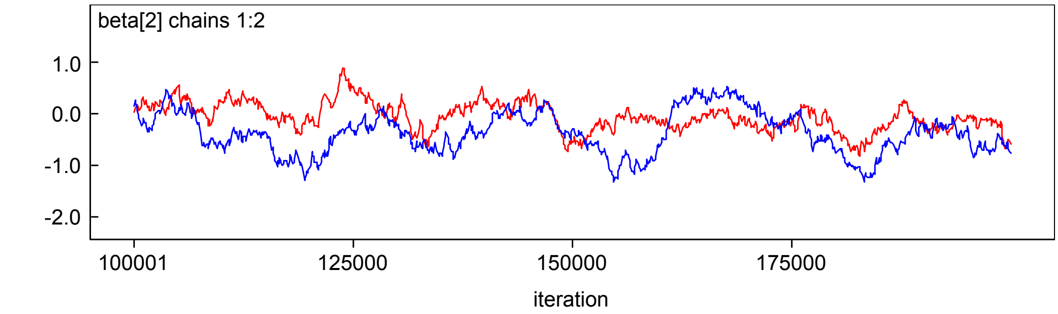


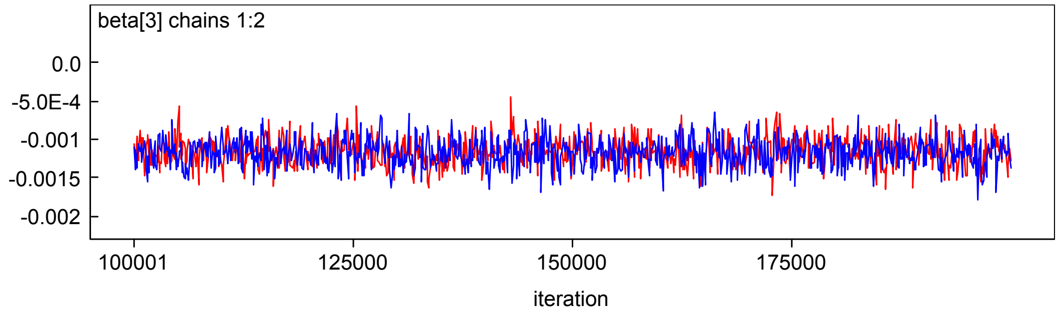
Appendix III: Kernel Density Graphs for parameters

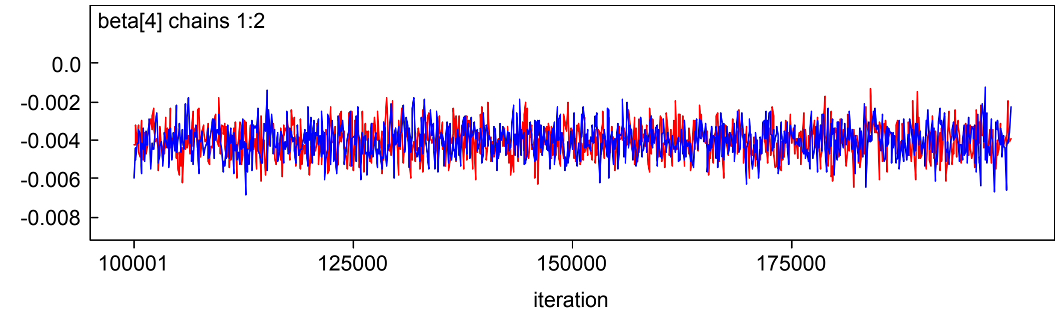


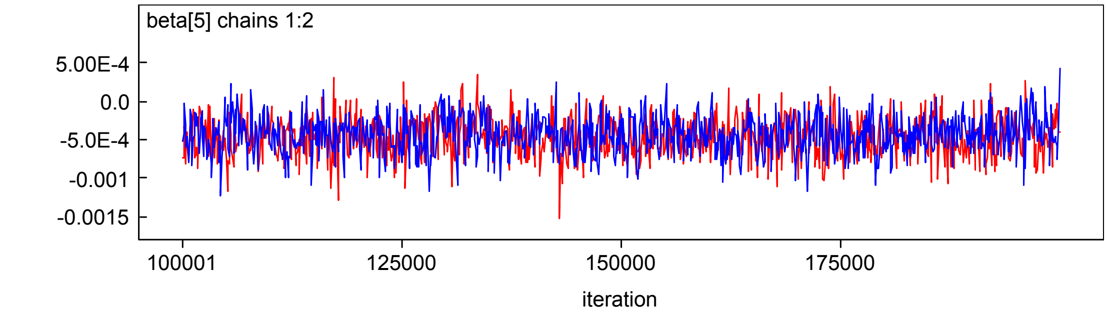
Appendix IV: Time series Graphs for parameters

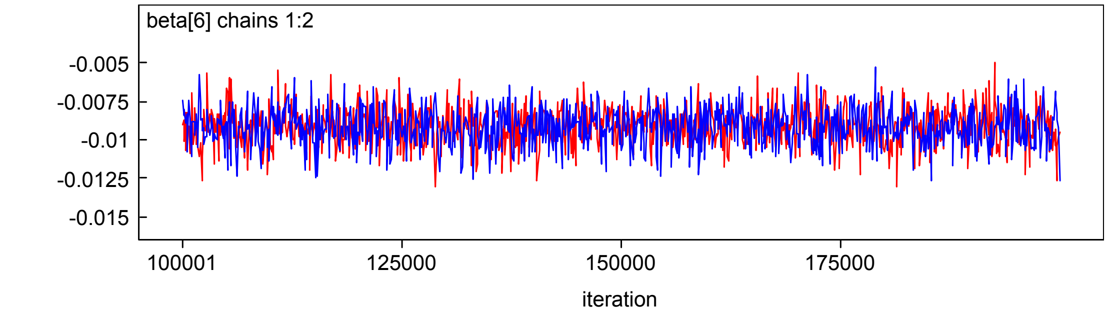


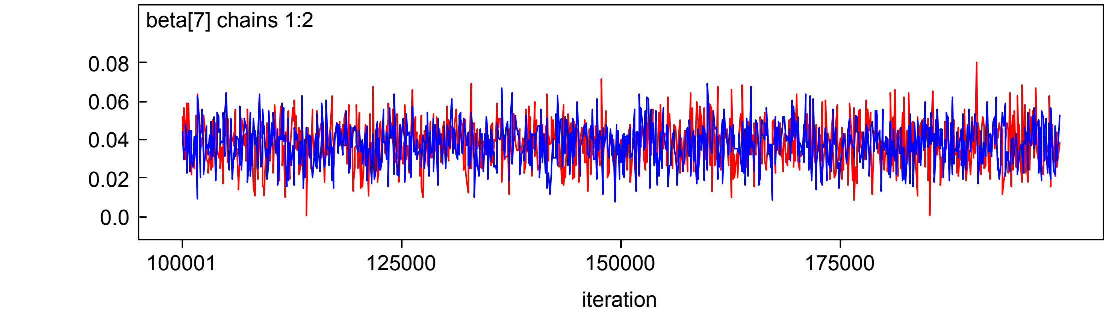


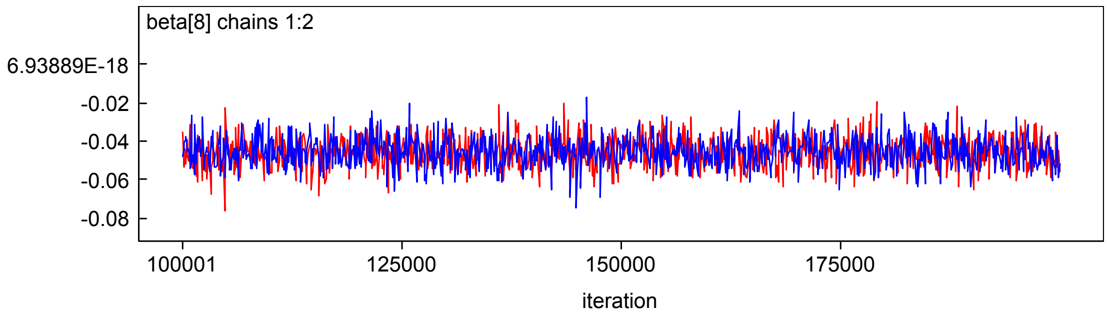


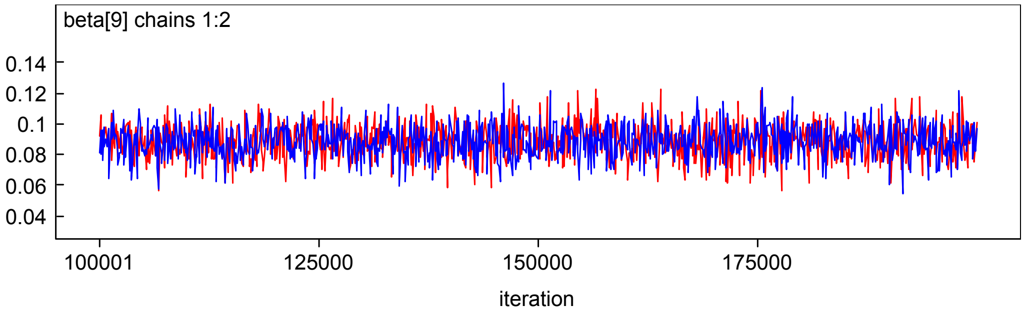


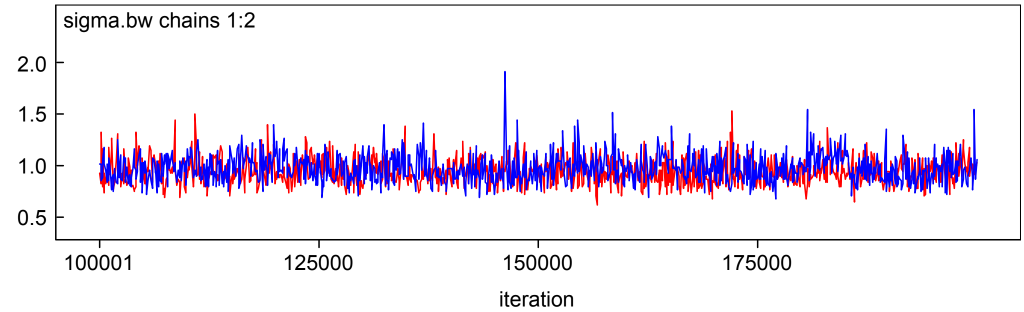

beta[5] chains 1:2
5.00E-4
0.0
-5.0E-4
-0.001
-0.0015
100001
125000
150000
175000
iteration

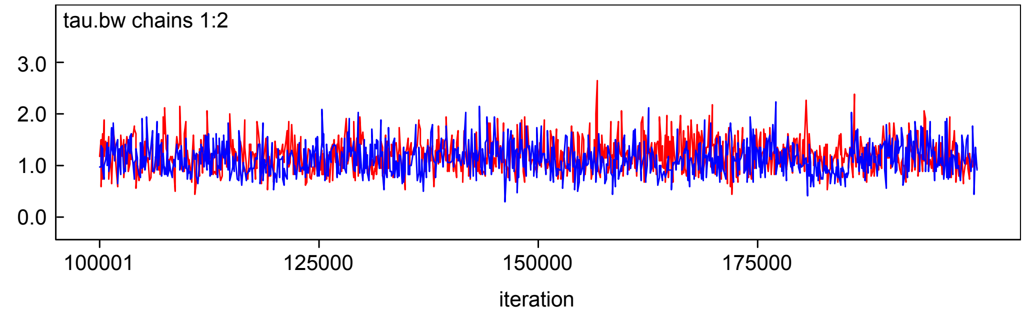

beta[6] chains 1:2
-0.005
-0.0075
-0.01
-0.0125
-0.015
100001
125000
150000
175000
iteration


beta[7] chains 1:2
0.08
0.06
0.04
0.02
0.0
100001
125000
150000
175000
iteration


beta[8] chains 1:2
6.93889E-18
-0.02
-0.04
-0.06
-0.08
100001
125000
150000
175000
iteration


beta[9] chains 1:2
0.14
0.12
0.1
0.08
0.06
0.04
100001
125000
150000
175000
iteration


sigma.bw chains 1:2
2.0
1.5
1.0
0.5
100001
125000
150000
175000
iteration


tau.bw chains 1:2
3.0
2.0
1.0
0.0
100001
125000
150000
175000
iteration

Appendix V: WinBUGS Code

```
model
{     for(j in 1 : N) {
           for(k in 1 : T[j]) {
               t[j,k] <- k
               log(lambda[j, k]) <-beta[1]+beta[2] * x1[j,k] + beta[3] * t[j,k]
               + beta[4] * (t[j,k] - t0[j])*step(t[j,k] - t0[j]) + beta[5]*x1[j,k]*t[j,k]
               + beta[6] * (x1[j,k] * (t[j,k] - t0[j])*step(t[j,k] - t0[j])) + beta[7] * x2[j,k]
               + beta[8] * x3[j,k] + beta[9] * x4[j,k]+  id[j]
               x2[j,k]<-cos(2*pi/4*s[j,k])
               x3[j,k]<-cos(2*pi/4*2*s[j,k])
               x4[j,k]<-sin(2*pi/4*s[j,k])
               mu[j,k] <- lambda[j,k] * x0[j,k] / 1000
                y[j, k] ~ dpois(mu[j,k]);
                                  }
           id[j]  ~ dnorm(0.0, tau.bw)
                      }
          for(i in 1:9) {beta[i] ~dnorm(0, 0.001)}
          tau.bw~dgamma(0.01,0.01)
          sigma.bw <- 1 / sqrt(tau.bw)
}

data
list(pi=3.1415926535,N=30,T=c(276,276,276,276,276,276,276,276,276,276,276,276,
276,276,276,276,276,276,276,252,276,276,276,276,276,276,276,276,276,276),
t0=c(138,257,235,235,235,250,198,198,210,220,212,210,211,222,222,138,257,211,2
35,235,250,198,198,210,220,212,210,211,222,222),
y=structure(.Data=c(...),.Dim=c(30,276)),
x0=structure(.Data=c(...),.Dim=c(30,276)),
x1=structure(.Data=c(...),.Dim=c(30,276)),
s=structure(.Data=c(...),.Dim=c(30,276)))

initials
list(beta=c(0,0,0,0,0,0,0,0,0),
tau.bw=1,id=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))
list(beta=c(0,-1,0,-1,0,0,0,0,0),
tau.bw=0.5,id=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))
```